People's Democratic Republic of Algeria.
Ministry of Higher Education and Scientific Research.
Mouloud Mammeri University of Tizi-Ouzou

Faculty of Sciences

Department of Mathematics

Bayesian inference in bivariate distribution models with applications

Thesis Submitted to the Department of Mathematics in Candidacy for Master Degree in Mathematics (Probability and Statistics)

Submitted by
Taous GALLEZE

Supervisor
Pr Hocine FELLAG (Tizi-Ouzou University, Algeria)

In collaboration with
Pr Juana Maria VIVO MOLINA (Murcia University, Spain)

Board of Examiners

Dr Lynda ATIL	UMMTO	Chairman
Pr Hocine FELLAG	UMMTO	Supervisor
Dr Cherifa BELKACEM	UMMTO	Member

Acknowledgments

First and foremost, I would like to express my deep and sincere commitment to my supervisor professor Hocine Fellag for his endless patience, suggestions and advices, as well as his support and encouragements throughout the development of this dissertation and during all my university curriculum.

I would like also to extend my gratitude to professor Juana Maria Vivo Molina for her invaluable guidance, vision, sincerity and motivation which have deeply inspired me. She was my second supervisor.

Special thanks are due to all the members of the jury including Ms.Atil Lynda and Ms.Belkacem Cherifa for having eagerly and willingly accepted to read and assess this work and for all the efforts they have provided so that we can have quality training.

Finally, I would like to express a special thanks to the team of Probability and Statistics and all the teachers of the department of mathematics of UMMTO.

Dedication

To my dear family
To my beloved fiance and his family
To my friends
To all of my teachers

Preface

This work is the result of a scientific collaboration between professors Hocine Fellag (Tizi-Ouzou, Algeria) and Juana Maria in Vivo (Murcia, Spain). Indeed, following the stay made by Professor Hocine Fellag at the University of Murcia where he carried out a teaching mobility for the benefit of master and doctoral students, the topic of this dissertation is proposed with a view to the preparation of a master's degree in mathematics, specializing in probability and statistics.

Contents

Introduction **1**

1 General notions on bivariate models **3**

1.1 Distributions of two-dimensional random variables . . . 3

1.1.1 Joint Cumulative Distribution Functions . . . 4

1.1.2 Joint Probability Density Functions . . . 5

1.1.3 Conditional Distribution, Density Functions . . . 7

1.2 Expected values involving a two-dimensional continuous random variables . 7

1.2.1 Definition . . . 8

1.2.2 Conditional Expected value . . . 8

1.2.3 Covariance . . . 9

1.2.4 Correlation coefficient . . . 9

1.2.5 Joints Moments . . . 9

1.3 Independent Random Variables . . . 9

1.4 Joint Characteristic Functions . . . 11

1.4.1 Joint characteristic function . . . 11

1.4.2 Joint probability generating function . . . 12

1.5 Transformations of a Two-Dimensional Random Variables . . . 12

1.5.1 Method 1, CDF Approach . . . 12

1.5.2 Method 2, Characteristic Function Approach . . . 12

1.5.3 Method 3, Conditional PDF Approach . . . 12

1.6 Bivariate Normal distribution . . . 13

2 Bivariate exponential distribution of the Marshall and Olkin **16**

2.1 Preliminary notions . . . 17

2.1.1 Survival function 17
2.1.2 Hazard function 17
2.1.3 Cumulative hazard function 17
2.1.4 Censoring data 18
2.1.5 Likelihood and Censoring 19
2.2 Univariate Exponential distribution 20
2.2.1 Definitions 20
2.2.2 Moments 20
2.2.3 Basic results and properties 23
2.2.4 Univariate extensions of the exponential distribution 26
2.2.5 Poisson Process 27
2.3 Marshall-Olkin's bivariate exponential distribution 28
2.3.1 Bivariate exponential models 28
2.3.2 Definition of Marshall-Olkin's BVE model 31
2.3.3 Derivation of the Marshall-Olkin's bivariate exponential distribution 32
2.3.4 Properties of Marshall-Olkin's bivariate exponential distribution . . 34
2.3.5 Comparison of MOBVE with independence case 40
2.3.6 Transformation of Marginals 41
2.3.7 Some studies on MOBVE model 42
2.3.8 Application fields 43
2.3.9 Comparison of the Marshall-Olkin distributions with Freund and Block and Basu distributions 44

3 Inference for Marshall-Olkin model 45
3.1 Preliminary notions in Bayesian paradigm 46
3.1.1 Bayes theorem 46
3.1.2 Definition 46
3.1.3 Loss function and risk 48
3.1.4 Usual loss functions 49
3.1.5 The choice of prior distributions 51
3.1.6 Credible intervals 53
3.1.7 MCMC methods 54
3.2 Classical inference for the Marshall-Olkin Model 57

3.3 Tests of independence 59
3.4 Bayesian inference for the Marshall-Olkin Model 61
3.4.1 Series sampling 61
3.4.2 Parallel sampling 63
3.4.3 Gain in precision of parallel over series sampling 64
3.5 Some recent works on Bayesian analysis of bivariate distributions 66
3.5.1 Objective Bayesian analysis for bivariate Marshall-Olkin exponential distribution 66
3.5.2 Marshall-Olkin bivariate Weibull distribution 66
3.5.3 Bayesian analysis of three parameter absolute continuous Marshall-Olkin bivariate Pareto distribution 68
3.5.4 Bayesian analysis of absolute continuous Marshall-Olkin bivariate Pareto distribution with location and scale parameters 70
3.5.5 Comparison of the maximum likelihood and Bayes estimators for symmetric bivariate exponential distribution under different loss function (Chadli et al, 2013) 70
3.6 Simulation study 72
3.6.1 Simulation of the bivariate Marshall-Olkin exponential distribution 72
3.6.2 Importance sampling 74
3.6.3 Numerical simulation 76
3.7 Real data application 78

Conculusion **80**

Bibliography **81**

Abbreviations

CDF: Cumulative Distribution Function.

PDF: Probability Distribution Function.

BVE: BiVariate Exponential distribution.

MOBVE: Marshall-Olkin BiVariate Exponential distribution.

MOBW: Marshall-Olkin Bivariate Weibull distribution.

MOBVPA: Marshall-Olkin BiVariate PAreto distribution.

BLMP: Bivariate Luck Memory Property.

Introduction

Because the importance of the univariate exponential distribution as a physical model, a variety of bivariate extensions of this distribution have been considered in the literature; see, for instance, Gumbel (1960), Freund (1961), Block and Basu (1974), Marshall and Olkin (1967) and so on.

Among these extensions, the bivariate Marshall-Olkin distribution (BVE) has received the maximum attention since this bivariate distribution has univariate exponential marginal distribution and a bivariate lack of memory property. Moreover, it has a nice physical interpretation based on random shocks. Consequently, as it can found in the literature, the Marshall-Olkin bivariate exponential distribution has widely been applied in reliability, finance and many more areas of science.

The main objective of the present dissertation is study the Bayesian estimation of the bivariate exponential distribution of Marshall and Olkin , and some of its generalizations available in the literature.

For that end, the specific objectives which will tackle are the following:

Revise the theoretical bivariate probability basis.

Revise the construction and the interpretation of the bivariate exponential distribution of the Marshall and Olkin.

Revise the distributional and stochastic properties of the Marshal-Olkin's bivariate exponential distribution.

Revise other well-known extensions of the univariate exponential distribution.

Compare the bivariate exponential distributions.

Revise the Bayesian estimation basis.

Revise the Bayesian estimation for Marshall-Olkin's bivariate exponential distribution and generalizations available in the literature.

Conduct a simulation study in order to illustrate the strengths of the Bayesian estimation.

Thus, this work is structured around three chapters. In the first one, we revise the theoretical bivariate probability notions. In the second chapter, firstly we recall some preliminary survival notions, the definition of the univariate exponential distribution and we give its properties. Then the definition of the Marshall-Olkin bivariate exponential distribution is given along with its stochastic properties and characterization. We recall also other well-known extensions of bivariate exponential distribution in order to compare them with Marshall-Olkin bivariate model.

In the third chapter, we discuss in detail, the problem of estimation, such a chapter is divided into three important sections. In the first section, we present some basic notions on Bayesian inference. In the second section, we study the Bayesian estimators of the unknown parameters for Marshall-Olkin exponential bivariate distribution and its generalizations available in the literature. Bayesian estimators are computed under proper priors. This section is closely related to the paper of Pena and Gupta (1990). Pena and Gupta (1990) obtained the Bayes estimators of the unknown parameters of the Marshall-Olkin bivariate model for both series and parallel systems under quadratic loss function. They have used a very flexible Dirichlet-Gamma conjugate prior, Jeffrey's noninformative prior can be obtained as a limiting case. In the last section, we prove that Bayesian approach performs better than the MLE's methodology. We illustrate this using a simulation study with importance sampling and using a real data study.

Chapter 1

General notions on bivariate models

Data in statistics are sometimes classified according to how many variables are in particular study. Depending on the number of variables being looked at, the data might be univariate, bivariate or it might be also multivariate.

• Univariate analysis is the simplest form of data analysis where the data being analyzed contains only one variable. Since it is a single variable it does not deal with causes or relationships.

• Bivariate analysis is used to find out if there is a relationship between two different variables (correlation, regression..).

• Multivariate analysis is used in the analysis of three or more variables.

In this chapter, we focus on bivariate analysis, where exactly two measurement are made on each observation.

1.1 Distributions of two-dimensional random variables

Definition

Let X and Y be two random variables on the probability space (Ω, A, P). A two-dimensional random variable (X,Y) is a mapping function (X,Y): $\Omega \to \mathbb{R}^2$, such that for any numbers $x, y \in \mathbb{R}$

$$\{\omega \in \Omega \mid X(\omega) \leq x \quad and \quad Y(\omega) \leq y\} \in A$$

1.1.1 Joint Cumulative Distribution Functions

Definition

The joint cumulative distribution function (CDF) of a two-dimensional random variables (X,Y), is given by

$$F_{XY}(x,y) = P(\{X \leq x\} \cap \{Y \leq y\}) = P(X \leq x, Y \leq y) \tag{1.1}$$

• The properties of joint CDF are summarized as follows:

1. $F_{XY}(-\infty,-\infty) = F_{XY}(-\infty,y) = F_{XY}(x,-\infty) = 0$;
2. $F_{XY}(\infty,\infty) = 1$;
3. $0 \leq F_{XY}(x,y) \leq 1$;
4. $F_{XY}(x,\infty) = F_X(x)$, $F_{XY}(\infty,y) = F_Y(y)$;
5. $P(x_1 < X_1 \leq x_2, y_1 < Y_1 \leq y_2) = F_{XY}(x_2,y_2) - F_{XY}(x_1,y_2) - F_{XY}(x_2,y_1) + F_{XY}(x_1,y_1) \geq 0$;

Property (5) tells us how to calculate the probability of two-dimensional random variables falling in a rectangular region.

Example

One of the simplest examples of a two-dimensional random variables is one that is uniformly distributed over the unit square

$$F_{XY}(x,y) = \begin{cases} 0 & x < 0 \ or \ y < 0 \\ x & 0 \leq x \leq 1, y > 1 \\ y & x > 1, 0 \leq y \leq 1 \\ xy & 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 1 & x > 1, y > 1 \end{cases}$$

Even this very simple example leads to a rather cumbersome function. Nevertheless, it is straightforward to verify that this function does indeed satisfy all the properties of a joint

CDF. From this joint CDF, the marginal CDF of X can be found to be

$$F_X(x) = F_{XY}(x, \infty) = \begin{cases} 0 & x < 0 \\ x & 0 \le x \le 1 \\ 1 & x > 1 \end{cases}$$

Hence, the marginal CDF of X is also a uniform distribution. The same statement holds for Y as well.

1.1.2 Joint Probability Density Functions

Definition

The joint probability density function (PDF) of a two-dimensional random variables (X,Y) evaluated at the point (x,y) is

$$f_{XY}(x, y) = \lim_{\epsilon_x \to 0, \epsilon_y \to 0} \frac{P(x \le X < x + \epsilon_x, y \le Y < y + \epsilon_y)}{\epsilon_x \epsilon_y} \tag{1.2}$$

Theorem

The joint PDF f_{XY} can be obtained from the joint CDF $F_{X,Y}$ by taking a partial derivative with respect to each variable. That is

$$\frac{\partial^2}{\partial x \partial y} F_{XY}(x, y) \tag{1.3}$$

Proof

Using the property (5) of CDF

$$\begin{aligned} P(x \le X < x + \epsilon_x \quad &, \quad y \le Y < y + \epsilon_y) \\ &= F_{XY}(x + \epsilon_x, y + \epsilon_y) - F_{XY}(x, y + \epsilon_y) - F_{XY}(x + \epsilon_x, y) + F_{XY}(x, y) \\ &= [F_{XY}(x + \epsilon_x, y + \epsilon_y) - F_{XY}(x, y + \epsilon_y)] - [F_{XY}(x + \epsilon_x, y) - F_{XY}(x, y)] \end{aligned}$$

Dividing by ϵ_x and taking the limit as $\epsilon_x \to 0$ results in

$$\begin{aligned}
\lim_{\epsilon_x \to 0} \frac{P(x \le X < x+\epsilon_x, y \le Y < y+\epsilon_y)}{\epsilon_x} &= \lim_{\epsilon_x \to 0} \frac{F_{XY}(x+\epsilon_x, y+\epsilon_y) - F_{XY}(x, y+\epsilon_y)}{\epsilon_x} \\
&- \lim_{\epsilon_x \to 0} \frac{F_{XY}(x+\epsilon_x, y) - F_{XY}(x, y)}{\epsilon_x} \\
&= \frac{\partial}{\partial x} F_{XY}(x, y+\epsilon_y) - \frac{\partial}{\partial x} F_{XY}(x, y)
\end{aligned}$$

Then dividing by ϵ_y and taking the limit as $\epsilon_y \to 0$ gives the desired result:

$$\begin{aligned}
f_{XY}(x,y) &= \lim_{\epsilon_x \to 0, \epsilon_y \to 0} \frac{P(x \le X < x+\epsilon_x, y \le Y < y+\epsilon_y)}{\epsilon_x \epsilon_y} \\
&= \lim_{\epsilon_y \to 0} \frac{\frac{\partial}{\partial x} F_{XY}(x, y+\epsilon_y) - \frac{\partial}{\partial x} F_{XY}(x, y)}{\epsilon_y} \\
&= \frac{\partial^2}{\partial x \partial y} F_{XY}(x, y)
\end{aligned}$$

This theorem shows that we can obtain a joint PDF from a joint CDF by differentiating with respect to each variable. The converse of this statement would be that we could obtain a joint CDF from a joint PDF by integrating with respect to each variable. Specifically,

$$F_{XY}(x,y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f_{XY}(u,v) du dv. \tag{1.4}$$

Properties

From the definition of the joint PDF in equation (1.2) as well as the relationships specified in equation (1.3) and (1.4), several properties of joint PDFs can be inferred. These properties are summarized as follows:

1. $f_{XY}(x,y) \ge 0$;
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x,y) dx dy = 1$;
3. $F_{XY}(x,y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f_{XY}(u,v) du dv$;
4. $f_{XY}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{XY}(x,y)$;

5. $f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)dy, \; f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)dx$

6. $P(x_1 < X_1 \leq x_2, y_1 < Y_1 \leq y_2) = \int_{y_1}^{y_2} \int_{x_1}^{x_2} f_{XY}(x,y)dxdy$

1.1.3 Conditional Distribution, Density Functions

The idea of this section is extended to the case where the conditioning event is related to another random variable.

Theorem

The conditional PDF of a random variable X given Y = y is

$$f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)}$$

Proof

Consider the conditioning event $A = y \leq Y < y + dy$. Then

$$\begin{aligned} f_{X|A}(x)dx &= P(x \leq X < x + dx | y \leq Y < y + dy) \\ &= \frac{P(x \leq X < x + dx, y \leq Y < y + dy)}{P(y \leq Y < y + dy)} \\ &= \frac{f_{XY}(x,y)dxdy}{f_Y(y)dy} = \frac{f_{XY}(x,y)dx}{f_Y(y)}. \end{aligned}$$

Passing to the limit as $dy \to 0$, the event A becomes the event Y =y, producing the desired result.

1.2 Expected values involving a two-dimensional continuous random variables

The notion of expected value is easily generalized to two-dimensional random variables. To begin, we define the expected value of an arbitrary function of two random variables.

1.2.1 Definition

Let g(x,y) be an arbitrary two-dimensional function. The expected value of g(X,Y), where X and Y are random variables, is

$$E[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{XY}(x,y)dxdy$$

If the function g(x,y) is actually a function of only a single variable, say x, then this definition reduces to the definition of expected values for functions of a single random variable

$$\begin{aligned} E[g(X)] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x) f_{XY}(x,y)dxdy \\ &= \int_{-\infty}^{\infty} g(x)\left[\int_{-\infty}^{\infty} f_{XY}(x,y)dy\right]dx \\ &= \int_{-\infty}^{\infty} g(x) f_X(x)dx \end{aligned}$$

To start with, consider an arbitrary linear function of the two variables g(x,y)= ax+by, where a and b are constants. Then

$$\begin{aligned} E[aX+bY] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} ax+by f_{XY}(x,y)dxdy \\ &= a\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{XY}(x,y)dxdy + b\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f_{XY}(x,y)dxdy \\ &= aE[X] + bE[Y] \end{aligned}$$

1.2.2 Conditional Expected value

Definition

The conditional expected value of a function g(X) of a random variable X given Y=y

$$E[g(X)|Y] = \int_{-\infty}^{\infty} g(x) f_{X|Y}(x|y)dx$$

1.2.3 Covariance

Definition

The covariance between two random variables is

$$Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)]$$

where μ_X and μ_Y are means of X and Y respectively.

1.2.4 Correlation coefficient

Definition

The correlation coefficient of two random variables X and Y, ρ_{XY}, is defined as

$$\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X\sigma_Y}$$

1.2.5 Joints Moments

Definition

The joint moments of two continuous random variables X and Y summarize information about their joint behavior. The $(j,k)^{th}$ moment of X and Y is defined by

$$E[X^jY^k] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^jy^kf_{XY}(x,y)dxdy$$

1.3 Independent Random Variables

Definition

Consider the events $A = \{X \leq x\}$ and $B = \{Y \leq y\}$ .The two events A and B are statistically independent if $P(A,B) = P(A)P(B)$.

Restated in terms of the random variables, this condition becomes

$$P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y) \Rightarrow F_{XY}(x, y) = F_X(x)F_Y(y)$$

Hence, two random variables are statistically independent if their joint CDF factors into a product of the marginal CDFs. Differentiating both sides of this equation with respect to both x and y reveals that the same statement applies to the PDF as well. That is, for statistically independent random variables, the joint PDF factors into a product of the marginal PDFs:

$$f_{XY}(x, y) = f_X(x)f_Y(y)$$

Theorem

Let X and Y be two independent random variables and consider forming two new random variables $U = g_1(X)$ and $V = g_2(Y)$. These new random variables U and V are also independent.

Proof

To show that U and V are independent, consider the events $A = (U \leq u)$ and $B = (V \leq v)$. Next, define the region R_u to be the set of all points x such that $g_1(x) = u$. Similarly, define R_v to be the set of all points y such that $g_2(y) = v$. Then

$$P(U \leq u, V \leq v) = P(X \in R_u, Y \in R_v) = \int_{R_u}\int_{R_v} f_{XY}(x, y)dxdy$$

Since X and Y are independent, their joint PDF can be factored into a product of marginal PDFs resulting in

$$\begin{aligned} P(U \leq u, V \leq v) &= \int_{R_u} f_X(x)dx \int_{R_v} f_Y(y)dy \\ &= P(X \in R_u)P(Y \in R_v) = P(U \leq u)P(V \leq v). \end{aligned}$$

Since we have shown that $F_{U,V}(u, v) = F_U(u)F_V(v)$, the random variables U and V must be independent.

Another important result deals with the correlation, covariance, and correlation coefficients of independent random variables.

Theorem

If X and Y are independent random variables, then

$$E[XY] = \mu_X \mu_Y,\ Cov(X,Y) = 0,\ \rho_{XY} = 0$$

Proof

$$E[XY] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf_{XY}(x,y) = \int_{-\infty}^{\infty} xf_X(x)dx. \int_{-\infty}^{\infty} yf_Y(y)dy = \mu_X \mu_Y$$

The conditions involving covariance and correlation coefficient follow directly from this result.

Remark

Independent random variables are necessarily uncorrelated, but the converse is not always true. Uncorrelated random variables do not have to be independent.

1.4 Joint Characteristic Functions

When computing the joint moments of random variables, it is often convenient to use characteristic functions. Since a pair of random variables is involved, the frequency domain function must now be two-dimensional.

1.4.1 Joint characteristic function

Definition

Given a two-dimensional random variables X and Y with a joint PDF, $f_{XY}(x,y)$, the joint characteristic function is

$$\Phi_{XY}(\omega_1, \omega_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{i(\omega_1 x + \omega_2 y)} f_{XY}(x,y)dxdy.$$

1.4.2 Joint probability generating function

1.5 Transformations of a Two-Dimensional Random Variables

Three methods have been developed for finding the PDF of $Z = g(X, Y)$ given the joint PDF of X and Y; they can be summarized as follows.

1.5.1 Method 1, CDF Approach

Define a set $R(z) = \{(x, y) : g(x, y) \leq z\}$. The CDF of Z is the integral of the joint PDF of X and Y over the region R(z). The PDF is then found by differentiating the expression for the CDF:

$$f_Z(z) = \frac{d}{dz} \int \int_{R(z)} f_{XY}(x, y) dx dy.$$

1.5.2 Method 2, Characteristic Function Approach

First, find the characteristic function of Z according to

$$\Phi_Z(\omega) = E\left[e^{i\omega g(X,Y)}\right].$$

Then compute the inverse transform to get the PDF of Z.

1.5.3 Method 3, Conditional PDF Approach

Fix either X = x or Y = y (whichever is more convenient). The conditional PDF of Z can then be found using the techniques developed for single random variables. Once the conditional PDF of Z is found, the unconditional PDF is given by

$$f_Z(z) = \int_{-\infty}^{\infty} f_{Z|Y}(z|y) f_Y(y) dy$$

or

$$f_Z(z) = \int_{-\infty}^{\infty} f_{Z|X}(z|x) f_X(x) dx$$

Next, our attention moves to solving a slightly more general class of problems. Given two random variables X and Y, now, suppose we create two new random variables W and Z according to some 2×2 transformation of the general form

$$Z = g_1(X,Y) \quad and \quad W = g_2(X,Y) \tag{1.7}$$

Then

$$f_{ZW}(z,w) = |J|^{-1} f_{XY}[h_1(z,w), h_2(z,w)]$$

with

$$|J| = det \begin{pmatrix} \frac{\partial h_1(z,w)}{\partial x} & \frac{\partial h_1(z,w)}{\partial y} \\ \frac{\partial h_2(z,w)}{\partial x} & \frac{\partial h_2(z,w)}{\partial y} \end{pmatrix}$$

with $X = h_1(z,w)$ and $Y = h_2(z,w)$ are the inverse transformation of equation (1.7).

1.6 Bivariate Normal distribution

Definition

A two-dimensional random variables X and Y are said to be jointly Gaussian if their joint PDF is of the general form

$$\begin{aligned} f_{XY}(x,y) &= \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{XY}^2}} \\ &\times \exp\left(-\left[\frac{(\frac{x-\mu_X}{\sigma_X})^2 - 2\rho_{XY}(\frac{x-\mu_X}{\sigma_X})(\frac{y-\mu_Y}{\sigma_Y}) + (\frac{y-\mu_Y}{\sigma_y})^2}{2(1-\rho_{XY}^2)}\right]\right) \end{aligned}$$

where μ_X and μ_Y are the means of X and Y, respectively; σ_X and σ_Y are the standard deviations of X and Y, respectively; and ρ_{XY} is the correlation coefficient of X and Y.

• The marginal of bivariate Gaussian distribution are Gaussian

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)dy = \frac{1}{\sqrt{2\pi\sigma_X^2}} \exp\left(-\frac{(x-\mu_X)^2}{2\sigma_X^2}\right)$$

and

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)dx = \frac{1}{\sqrt{2\pi\sigma_Y^2}} \exp\left(-\frac{(y-\mu_Y)^2}{2\sigma_Y^2}\right)$$

• The conditional function of X given Y=y is also Gaussian

$$f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)} = \frac{1}{\sqrt{2\pi\sigma_X}\sqrt{1-\rho_{XY}^2}} \exp\left\{\frac{-1}{2\sigma_X^2(1-\rho_{XY}^2)}\left(x - \mu_X \frac{\rho_{XY}\sigma_X}{\sigma_Y(y-\mu_Y)}\right)^2\right\}$$

Theorem

If the correlation between two Gaussian random variables equal to zero, the variables are independent.

Proof

Uncorrelated Gaussian random variables have a correlation coefficient of zero. Plugging $\rho_{XY} = 0$ into the general joint Gaussian PDF results in

$$f_{XY}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y} \exp\left(-\left[\frac{(\frac{x-\mu_X}{\sigma_X})^2 + (\frac{y-\mu_Y}{\sigma_y})^2}{2}\right]\right)$$

This clearly factors into the product of the marginal Gaussian PDFs

$$f_{XY}(x,y) = \frac{1}{\sqrt{2\pi\sigma_X^2}} \exp\left(-\frac{(x-\mu_X)^2}{2\sigma_X^2}\right) \frac{1}{\sqrt{2\pi\sigma_Y^2}} \exp\left(-\frac{(y-\mu_Y)^2}{2\sigma_Y^2}\right) = f_X(x)f_Y(y)$$

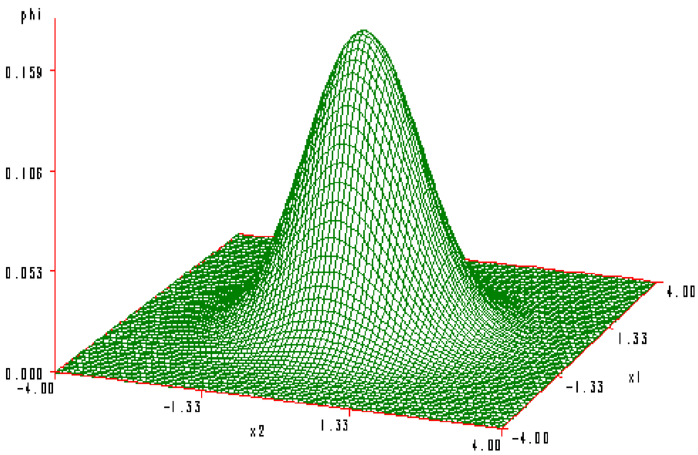

Figure 1.1: Bivariate normal distribution with zero correlation

Chapter 2

Bivariate exponential distribution of the Marshall and Olkin

The univariate exponential distribution is very popular among researchers working in the areas such as reliability analysis, life testing, and survival analysis. When there are two or more variables affecting the system, in most of the cases the analysis is carried out by assuming that they are statistically independent. However, the assumption of independence does not hold in practice. As a consequence, several bivariate models have been introduced in the literature. A detailed up-to-date survey of bivariate distributions and the methods of their constructions along with their applications may be found in book by Balakrishnan and Lai (2009). Such distributions are important candidates for joint modeling of non-negative variables such as lifetime and repair time of equipments. Some well known bivariate exponential distributions are those by Gumbel (1960), Freund (1961), Marshall and Olkin (1967), Block and Basu (1974) and so on.
In our work, we confine ourselves to the Marshall-Olkin's bivariate exponential distribution denoted MOBVE. But before starting by defining this model and given its properties, it is necessary to recall some preliminary survival notions.

2.1 Preliminary notions

2.1.1 Survival function

Let T denote the time from a well-defined starting point until some event called failure, occurs. T is referred to as survival time, let f(t) denote its probability density function and F(t) its cumulative distribution function.
The probability that the survival time T exceeds some value t, S(t), is given by:

$$S(t) = P(T > t) = \int_t^{\infty} f(t)dt = 1 - F(t) = \bar{F}(t) \tag{2.1}$$

•S(t) is a non-increasing function,
• S(0)=1 and $S(+\infty) = 0$.

2.1.2 Hazard function

A quantity that is often used along the survival function is the hazard function which given by

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t < T < t + \Delta t | T \geq t)}{\Delta t} = \frac{f(t)}{S(t)} \tag{2.2}$$

Note that you can also write the hazard function

$$h(t) = -\frac{\partial \log S(t)/\partial t}{S(t)} \tag{2.3}$$

• The hazard function describes the "intensity of death" at the time t given that the individual has already survived past time t.

2.1.3 Cumulative hazard function

There is an other quantity that is also common in survival analysis, the cumulative hazard function which is given by

$$H(t) = \int_0^t h(s)ds \tag{2.4}$$

You can interpret H(t) as the cumulative amount of hazard up to time t.
The cumulative hazard function and survival function as linked as follows:

$$H(t) = -\log S(t) \quad (2.5),$$

$$S(t) = e^{-H(t)} = e^{-\int_0^t h(s)ds} \quad (2.6)$$

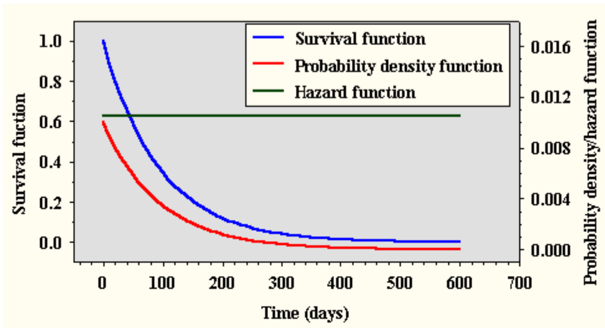

Figure 2.1: Survival function, probability density function and hazard function for population with consatant hazard function

2.1.4 Censoring data

Censoring in a study is when there is incomplete information about a study participant, observation or value of measurement. In clinical trials, it is when the event does not happen while the subject is being monitored or because they drop out of the trial. There are three types of censoring data:

Right censoring

Right censoring occurs when a subject leaves the study before an event occurs, or the study ends before the event has occurred. For example, we consider patients in a clinical trial to study the effect of treatments on stroke occurrence. The study ends after 5 years. Those patients who have had no strokes by the end of the year are censored. If the patient leaves the study at time t_e; then the event occurs in $(t_e; \infty)$. Two types of independent right censoring:

Type I: completely random dropout (eg emigration) and/or fixed time of end of study no

event having occurred.
Type II: study ends when a fixed number of events amongst the subjects has occurred.

Left censoring

Left censoring is when the event of interest has already occurred before enrolment. This is very rarely encountered.

Interval censoring

By interval censoring, we mean that a random variable of interest in known only to lie within an interval instead of being observed exactly. For example, if we do not know exactly when some babies are born but we know it is within some interval of time.

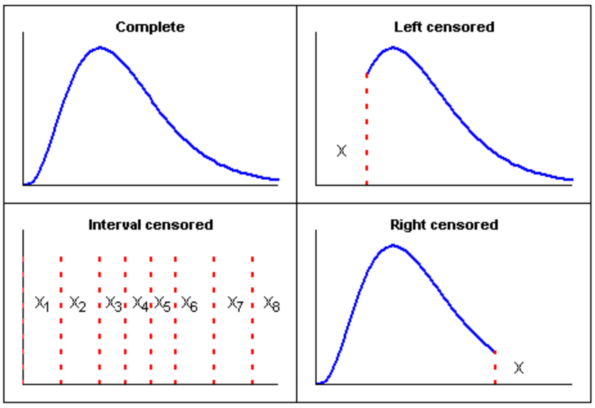

Figure 2.2: Types of censoring data

2.1.5 Likelihood and Censoring

If the censoring mechanism is independent of the event process, then we have an easy way of dealing with it.
Suppose that T is the time to event and that C is the time to the censoring event.
Assume that all subjects may have an event or be censored, say for subject i one of a pair of observations $(\tilde{t}_i, \tilde{c}_i)$ may be observed: Then since we observe the minimum time we would have the following expression for the likelihood (using independence)

$$L = \prod_{\tilde{t}_i < \tilde{c}_i} f(\tilde{t}_i) S_C(\tilde{t}_i) \prod_{\tilde{c}_i < \tilde{t}_i} S(\tilde{c}_i) f_C(\tilde{c}_i)$$

Now define the following random variable:

$$\delta = \begin{cases} 1 & T < C \\ 0 & T > C \end{cases}$$

For each subject we observe $t_i = \min(\tilde{t}_i, \tilde{c}_i)$ and δ_i; observations from a continuous random variable and a binary random variable. In terms of these L become

$$L = \prod_i h(t_i)^{\delta_i} S(t_i) \prod_i h_C(t_i)^{1-\delta_i} S_C(t_i)$$

where we have used density = hazard × survival function

2.2 Univariate Exponential distribution

2.2.1 Definitions

• A continuous random variable X is said to have an exponential distribution with parameter $\lambda > 0$, shown as X $\sim$ Exponential(λ) if its probability density function (pdf) is given by

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases} \tag{2.7}$$

• The cumulative distribution function is given by

$$F_X(x;\lambda) = \int_{\infty}^{x} f(t)dt = \int_{0}^{x} f(t)dt = \begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases} \tag{2.8}$$

2.2.2 Moments

The r-th moment of a random variable X is given by

$$E[X^r] = \int_0^{\infty} x^r f(x)dx \tag{2.9}$$

Thus when X is exponentially distributed,

$$E[X^r] = \int_0^\infty x^r \lambda e^{-\lambda x} dx$$

Let $\mu = \lambda x$, $dx = \frac{1}{\lambda} d\mu$

$$\begin{aligned} E[X^r] &= \lambda \int_0^\infty (\frac{\mu}{\lambda})^r e^{-\mu} \frac{1}{\lambda} d\mu = \frac{1}{\lambda^r} \int_0^\infty \mu^r e^{-\mu} d\mu \\ &= \frac{1}{\lambda^r} \Gamma(r+1) = \frac{r!}{\lambda^r} \end{aligned}$$

for positive integer r.

Mean

The mean

$$\mu = E[X] = \frac{1}{\lambda}$$

Variance

The variance

$$\sigma^2 = E[X^2] - \mu^2 = \frac{1}{\lambda^2}$$

Skewness

The skewness of a curve γ_1 , is a measure of the symmetry of the curve in comparison with the normal curve and is given by

$$\gamma_1 = \frac{\mu_3}{\sigma^3}$$

with $\mu_n = E[X-\mu]^n$
Thus for an exponential curve,

$$\gamma_1 = \frac{\frac{2}{\lambda^3}}{(\frac{1}{\lambda})^3} = 2$$

The normal curve has $\gamma_1 = 0$

Kurtosis

The kurtosis of a curve, γ_2 , is a measure of the sharpness of its peak and the width and length of its tail in comparison with the normal curve and is given by

$$\gamma_1 = \frac{\mu_4}{\sigma^4}$$

Thus for an exponential curve,

$$\gamma_2 = \frac{\frac{9}{\lambda^4}}{(\frac{1}{\lambda})^4} = 9$$

The normal curve has $\gamma_2 = 3$

Moment generating function

The moment generating function is

$$M(t) = E[e^{tx}] = \int_0^\infty e^{tx}\lambda e^{-\lambda x}dx = \frac{\lambda}{\lambda - t}$$

Cumulant generating function

The cumulate generating function is

$$\varphi(t) = E[e^{itx}] = \frac{\lambda}{\lambda - it}$$

Survival function

$$S(t) = 1 - F(t) = e^{-\lambda t}$$

Hazard function

The hazard rate function is

$$h(t) = \frac{f(t)}{S(t)} = \frac{\lambda e^{-\lambda x}}{e^{-\lambda x}} = \lambda, a\ constant$$

2.2.3 Basic results and properties

Sums of independent exponential random variables

Let $S_N = X_1 + X_2 + ... + X_N$
The distribution of S_N is Gamma with parameters n and λ.

Proof

let us start with

$$\begin{aligned} f_{X_1+X_2}(t) &= \int_0^t f_{X_1}(s).f_{X_2}(t-s) = \int_0^t \lambda e^{-\lambda s}\lambda e^{-\lambda(t-s)} \\ &= \lambda^2 \int_0^t e^{-\lambda s-\lambda t+\lambda s} = \lambda^2 e^{-\lambda t}t = \lambda e^{-\lambda t}(\lambda t) \end{aligned}$$

Next

$$\begin{aligned} f_{X_1+X_2+X_3}(t) &= \int_0^t f_{X_1+X_2}(s).f_{X_3}(t-s) = \int_0^t \lambda e^{-\lambda s}\lambda s.\lambda e^{-\lambda(t-s)} \\ &= \lambda^2.\lambda \int_0^t s e^{-\lambda s-\lambda t+\lambda s} = \lambda^2\lambda \int_0^t s e^{-\lambda t} \\ &= \lambda^2\lambda e^{-\lambda t}\left[\frac{s^2}{2}\right]_0^t = \lambda e^{-\lambda t}\frac{(\lambda t)^2}{2!} \end{aligned}$$

By induction, assume that $X_1 + X_2 + ... + X_{n-1}$ has a pdf given by

$$f_{X_1+X_2+...+X_{n-1}} = \lambda e^{-\lambda t}\frac{(\lambda t)^{n-2}}{(n-2)!}$$

Hence

$$\begin{aligned}
f_{X_1+X_2+...+X_n}(t) &= \int_0^t f_{X_1+X_2+...+X_{n-1}}(s) + f_{X_n}(t-s) = \int_0^t \lambda e^{-\lambda s}\frac{(\lambda s)^{n-2}}{(n-2)!}\lambda e^{-\lambda(t-s)} \\
&= \frac{\lambda^n e^-\lambda t}{(n-2)!}\int_0^t s^{n-2} = \frac{\lambda^n e^-\lambda t}{(n-2)!}\left[\frac{s^n-1}{n-1}\right]_0^t \\
&= \lambda e^{-\lambda t}\frac{(\lambda t)^{n-1}}{(n-1)!} = \frac{\lambda^n t^{n-1}}{\Gamma(n)}e^{-\lambda t}, t>0 \\
&= \frac{\lambda^n}{\Gamma(n)}e^{-\lambda t}t^{n-1}, t>0
\end{aligned}$$

which is a Gamma distribution with parameters n and λ

Distribution of the minimum of exponential random variables

Let $X_1, ..., X_n$ be independent exponentially distributed random variable with rate parameters $\lambda_1, ..., \lambda_n$. Then $T = \min(X_1, ..., X_n)$ is also exponentially distributed with parameter $\lambda = \sum_{i=1}^n \lambda_i$

Proof

$$\begin{aligned}
\bar{F}(t) &= P(\min_{1\le i\le n} X_i > x) \\
&= P(X_1 > x, ..., X_n > x) \\
&= \prod_{i=1}^n P(X_i > x) \\
&= \exp\{-(\sum_{i=1}^n \lambda_i)x\} = \exp(-\lambda x)
\end{aligned}$$

Memoryless property

Definition

A random variable X is said to be without memory or memory-less if

$$\bar{F}(t+x|x) = \frac{\bar{F}(t+x)}{\bar{F}(x)} = \bar{F}(t)$$

or equivalently

$$\bar{F}(t+x) = \bar{F}(t)\bar{F}(x) \ for \ all \ x, t \geq 0 \tag{2.9}$$

Then we have the following theorem

Theorem

The exponential distribution is the only continuous distribution which is memory-less

Proof

First we assume that (2.9) holds. Now let c>0, and let m and n be positive integers. Using (2.9) we get

$$\bar{F}(nc) = (\bar{F}(c))^n, \tag{2.10}$$

and

$$\bar{F}(c) = (\bar{F}(\frac{c}{m}))^m \tag{2.11}$$

Putting $c = \frac{1}{m}$ in (2.10) and c =1 in (2.11) we obtain

$$\bar{F}(\frac{n}{m}) = (\bar{F}(\frac{1}{m}))^n = (\bar{F}(1))^{n/m} \tag{2.12}$$

We claim that $0 < \bar{F}(1) < 1$. Because if $\bar{F}(1) = 1$, then by substituting c=1 in (2.10), $\bar{F}(n) = \bar{F}^n(1) = 1$ for all positive integers n which contradicts $\bar{F}(+\infty) = 0$
On the other hand if $\bar{F}(1) = 0$, then by (2.11) $\bar{F}(\frac{1}{m}) = (\bar{F}(1))^{1/m} = 0$, and hence using right continuity of $\bar{F}$, $\bar{F}(0) = 0$ which contradicts the non-degeneracy of T.
Thus $0 < \bar{F}(1) < 1$. Write $\bar{F}(1) = \exp(-\lambda)$, $0 < \lambda < \infty$.
Then it is clear that $\bar{F}(\frac{n}{m}) = \exp(-\lambda\frac{n}{m})$ for all positive rationales $\frac{n}{m}$.
By right continuity it follows that $\bar{F}(t) = \exp(-\lambda t)$ for all $t > 0$.
Using right continuity one more time we get $\bar{F}(0) = 1$. Now if $\bar{F}(t) = \exp(-\lambda t)$, then it is clear that $\bar{F}(t+x) = \bar{F}(t)\bar{F}(x) \ for \ all \ t, x > 0$
This completes the proof.

2.2.4 Univariate extensions of the exponential distribution

In statistics, they exist many distributions which are based on exponential distribution for example Gamma distribution and Weibull distribution.

Gamma distribution

It is often used in Bayesian statistics as conjugate prior distribution for exponential distribution and Poisson distribution.

Definition

A continuous random variable X is said to have a Gamma distribution with parameters $\alpha > 0$ and $\lambda > 0$, shown as $X \sim Gamma(\alpha, \lambda)$, if its PDF is given by

$$f(x|\alpha, \lambda) = \begin{cases} \frac{\lambda^{\alpha} x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)} & x > 0 \\ 0 & \text{otherwise} \end{cases}$$

when $\alpha = 1$, this becomes the exponential distribution.

Weibull distribution

It is the most popular distribution which generalizes the exponential distribution.

Definition

A continuous random variable X is said to have a Weibull distribution with parameters $\alpha > 0$ and $\lambda > 0$, shown as $X \sim W(\alpha, \lambda)$, if its PDF is given by

$$f(x|\alpha, \lambda) = \begin{cases} \frac{\alpha}{\lambda}(\frac{x}{\lambda})^{\alpha-1} e^{-(\frac{x}{\lambda})^{\alpha}} & x > 0 \\ 0 & \text{otherwise} \end{cases}$$

•The CDF of Weibull distribution is given by

$$F(x|\alpha, \lambda) = \begin{cases} 1 - e^{-(\frac{x}{\lambda})^{\alpha}} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

•Its survival function is

$$\bar{F}(x|\alpha,\lambda) = \begin{cases} e^{-(\frac{x}{\lambda})^{\alpha}} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

when $\alpha = 1$, this becomes the exponential distribution.

2.2.5 Poisson Process

In practice, the exponential distribution represents the probability distribution of the time between events in a Poisson point process.

Definition

A counting process is an integer valued process $\{N(t); t \geq 0\}$ which counts the number of events occurring in an interval following a random mechanism, so that the time to occurrence of each event is a random variable. M(t) = E[N(t)] is the expected number of events in [0,t] and $M'(t)$, if is exists, it called the intensity or mean rate.

Definition

A counting process $\{N(t); t \geq 0\}$ is a Poisson process with mean rate (or intensity) λ, if the following assumptions are fulfilled:

1. N(0) = 0
2. $\{N(t); t \geq 0\}$ has stationary independent increments. That is, for all choices of indices $t_0 < t_1 < ... < t_n$, the n random variables $N(t_i) - N(t_{i-1})$, i = 1, ... , n, are independent and $N(s_2+h)-N(s_1+h)$ and $N(s_2)-N(s_1)$ have the same distribution for all choices of indices s_1, s_2 and for every h > 0;
3. For any pair of times s, t such that s < t, the number of counts of occurrence of an event of N(t)- N(s) in the interval [s, t] has a Poisson distribution with mean $\lambda(t-s)$.

$$P(N(t) - N(s) = k) = \exp\{-\lambda(t-s)\}\frac{\{\lambda(t-s)\}^k}{k!}, k = 0, 1, 2....$$

Theorem

Let $\{N(t); t \geq 0\}$ be a Poisson process with mean rate λ. Then the length of the interval from some fixed time to the next event has the probability density function $f(x) = \lambda \exp(-\lambda x)$

Proof

Let T denote the time to occurrence of the first event from some fixed time T_0 . Then,

$$\begin{aligned} P(T > t) &= P(first\ event\ from\ T_0\ occurs\ after\ t) \\ &= P(N(T_0 + t) - N(T_0)) = 0 \\ &= \exp(-\lambda x) \end{aligned}$$

The last equality comes from (3) of the previous definition.

2.3 Marshall-Olkin's bivariate exponential distribution

Before studying the Marshall-Olkin's bivariate distribution, it is interesting to define some other bivariate exponential model such as the model introducing by Gumbel (1960), Freund (1961) and Block and Basu (1974).

2.3.1 Bivariate exponential models

Gumbel's Distributions

Gumbel (1960) is one of the first to consider bivariate exponential distributions. He proposed three, which are absolutely continuous models. None of these are physically motivated. We consider them below.

Model 1: The joint distribution of (X, Y) is given by

$$F(x, y) = 1 - e^{-x} - e^{-y} + e^{-x-y-\delta xy}, \quad x, y > 0, 0 < \delta < 1$$

Here, X and Y are independent when $\delta = 0$. Also, the marginal distributions are exponential distributions. Here, the joint density function is

$$f(x, y) = \exp\{-(x + y + \delta xy)\}\{(1 + \delta x)(1 + \delta y) - \delta\}$$

Model 2: Gumbel's second model is given by

$$F(x, y) = (1 - e^{-x})(1 - e^{-y})(1 + \alpha e^{-x-y}), \quad x, y > 0, -1 < \alpha < 1$$

Here, the joint density is given by

$$f(x, y) = e^{-x-y}\{1 + \alpha(2e^{-x} - 1)(2e^{-y} - 1)\}$$

X and Y are independent when $\alpha = 0$.

Model 3: Gumbel's third model is given by distribution function

$$F(x, y) = 1 - e^{-x} - e^{-y} + \exp\{-(x^m + y^m)^{1/m}\}, \quad x, y > 0, m > 1$$

and density function

$$f(x, y) = \exp\{-(x^m + y^m)^{1/m}\}(x^m + y^m)^{1/m-2}x^{m-1}y^{m-1}\{(x^m + y^m)^{1/m} + m - 1\}$$

Here, m = 1 corresponds to the case of independence.

Freund's Model

Freund (1961) is one of the first to consider a model that is physically motivated. Let X and Y denote the lifetimes of components A and B of a two-component system. To start with, X and Y are independent with $X \sim \exp(\alpha)$ and $Y \sim \exp(\beta)$. If A fails first, the lifetime of B is changed to $\exp(\beta')$ because of extra stress. If B fails first, the distribution of A changes to $\exp(\alpha')$. Assume $\alpha + \beta - \beta' \neq 0$. The joint density of (X, Y), under the above assumption, is given by

$$f(x, y) = \begin{cases} \alpha\beta' \exp[-\beta' y - (\alpha + \beta - \beta')x] & 0 < x < y \\ \alpha'\beta \exp[-\alpha' x - (\alpha + \beta - \alpha')y] & 0 < y < x \end{cases}$$

However, the marginal distributions of X and Y are not univariate exponential distributions in general. These are mixtures or weighted averages of exponential distributions and are given by

$$f(x) = \frac{(\alpha - \alpha')(\alpha + \beta)\exp(-(\alpha + \beta)x)}{\alpha + \beta - \alpha'} + \frac{\alpha'\beta\exp(-\alpha' x)}{\alpha + \beta - \alpha'}$$

provided $\alpha + \beta - \alpha' \neq 0$, and

$$g(y) = \frac{(\beta - \beta')(\alpha + \beta)\exp(-(\alpha + \beta)y)}{\alpha + \beta - \alpha'} + \frac{\alpha\beta'\exp(-\beta' y)}{\alpha + \beta - \alpha'}$$

Here, X and Y are independent if and only if $\alpha = \alpha'$ and $\beta = \beta'$. Note if $\beta > \beta'$ or $(\alpha > \alpha')$, the expected life of component B(A) improves when component A(B) fails. In this case,$\alpha + \beta - \alpha' > 0$ and the marginal densities are mixtures of exponential densities. On the other hand, if $\beta < \beta'$ $(\alpha < \alpha')$ the expected lifetime of B(A) decreases after the failure of component A(B). In this case, $\alpha + \beta - \alpha' < 0$ $(\alpha + \beta - \alpha' < 0)$ and the marginal densities are weighted averages of exponential densities.

Block-Basu Model

The model proposed by Block and Basu (1974) is closely related to the MOBVE and the model proposed by Freund. Let us call this the ACBVE. The MOBVE has univariate exponential marginals, and it satisfies the BLMP. However, it is not absolutely continuous. Block and Basu (1974) showed that it is not possible to have an absolutely continuous bivariate distribution that satisfies BLMP and has exponential marginal distributions. Since absolute continuity is a desirable criterion, Block and Basu proposed the ACBVE where the marginals are mixtures or weighted averages of exponential distributions. Here, the joint survival function is given by

$$\begin{aligned}\bar{F}(x,y) &= \exp[-\lambda_1 x - \lambda_2 y - \lambda_{12}\max(x,y)] \\ &- (\frac{\lambda_{12}}{\lambda_1 + \lambda_2}\exp[-\lambda \max(x,y)], \quad x, y > 0, \lambda_1, \lambda_2, \lambda_{12} > 0\end{aligned}$$

The joint density is given by

$$f(x,y) = \begin{cases} \frac{\lambda_1\lambda(\lambda_2+\lambda_{12})}{\lambda_1+\lambda_2}\exp[-\lambda_1 x - (\lambda_2 + \lambda_{12})y] & 0 < x < y \\ \frac{\lambda_2\lambda(\lambda_1+\lambda_{12})}{\lambda_1+\lambda_1}\exp[-(\lambda_1 + \lambda_{12})x - \lambda_2 y] & 0 < y < x \end{cases}$$

Here, the marginal survival functions are given by

$$\bar{F}_1(x) = \frac{\lambda}{\lambda_1 + \lambda_2} \exp[-(\lambda_1 + \lambda_{12})x] - \frac{\lambda_{12}}{\lambda_1 + \lambda_2} \exp(-\lambda)$$

$$\bar{F}_2(y) = \frac{\lambda}{\lambda_1 + \lambda_2} \exp[-(\lambda_2 + \lambda_{12})y] - \frac{\lambda_{12}}{\lambda_1 + \lambda_2} \exp(-\lambda)$$

Note the ACBVE is the absolutely continuous component of the BVE. However, it is not a special case, since the MOBVE has exponential marginals and the ACBVE does not. The ACBVE is a special case of the Freund model as it can be derived in the same way with special choice of the parameters. Assume $\alpha < \alpha'$ and $\beta < \beta'$ in Freund's model and let

$$\alpha = \lambda_1 + \lambda_{12}(\lambda_1/(\lambda_1 + \lambda_{12})), \quad \alpha' = \lambda_1 + \lambda_{12}$$

and

$$\beta = \lambda_2 + \lambda_{12}(\lambda_2/(\lambda_2 + \lambda_{12})), \quad \beta' = \lambda_2 + \lambda_{12}$$

Substituting these values of a α, α', β and β' in the joint density of Freund's model we obtain the joint density of Block's model.

The ACBVE, like BVE, has the BLMP and $\min(x, y) \sim \exp(\lambda)$ Also, $\lambda_{12} = 0$ corresponds to the case when X and Y are independent. However, unlike the BVE, the ACBVE is absolutely continuous. Other properties of the ACBVE are given in Block and Basu (1974).

2.3.2 Definition of Marshall-Olkin's BVE model

The bivariate exponential and the multivariate extension of exponential distributions due to Albert W. Marshall and Ingram Olkin (1967) has received considerable attention in describing the statistical dependence of components in a two component system developing statistical inference procedures.

The joint survival function for the Marshall-Olkin model (1967) is given by

$$\bar{F}(s,t) = P(X > s, Y > t) = \exp[-\lambda_1 s - \lambda_2 t - \lambda_{12}\max(s,t)] \quad s,t > 0 \tag{2.13}$$

that is,

$$\bar{F}(s,t) = \begin{cases} \exp\{-\lambda_1 s - (\lambda_2+\lambda_{12})t\} & 0 \le s \le t \\ \exp\{-(\lambda_1+\lambda_{12})s - \lambda_2 t\} & 0 \le t \le s \end{cases} \tag{2.14}$$

2.3.3 Derivation of the Marshall-Olkin's bivariate exponential distribution

This distribution arises in the following context: suppose there is a two-component system subject to shocks, which may knock out the first component, the second component, or both of them. If these shocks result from independent Poisson processes with parameters λ_1, λ_2 and λ_{12}, respectively, and if X and Y are the lifetimes of the first and the second components then their survival function is (2.13).

A fatal shock model

Suppose that the components of two-component system die after receiving a shock which is fatal. Independent Poisson process $Z_1(t;\lambda_1)$, $Z_2(t;\lambda_2)$ and $Z_{12}(t;\lambda_{12})$ govern the occurrence of shocks.
•Events in the process $Z_1(t;\lambda_1)$ are shocks to component 1.
•Events in the process $Z_2(t;\lambda_2)$ are shocks to component 2.
•Events in the process $Z_{12}(t;\lambda_{12})$ are shocks to both component .
Thus if X and Y denote the life of the first and second components,

$$\begin{aligned} \bar{F}(s,t) &\equiv P\{X > s, Y > t\} \\ &= P\{Z_1(s;\lambda_1) = 0, Z_2(t;\lambda_2) = 0, Z_{12}(\max(s,t);\lambda_{12}) = 0\} \\ &= \exp[-\lambda_1 s - \lambda_2 t - \lambda_{12}\max(s,t)]. \end{aligned}$$

Non-fatal shock models

Again consider a two-component system and three independent Poisson process $Z_1(t;\delta_1)$, $Z_2(t;\delta_2)$, $Z_{12}(t;\delta_{12})$ governing the occurrence of shocks to first component, second component and both components respectively. The shock need not to be fatal.
Describe the state of the system by the ordered pairs (0,0),(0,1),(1,0),(1,1), where a 1 in

the first (second) indicates that the first (second) component is operating and a 0 indicates that it is not.

•Events in the process $Z_1(t;\delta_1)$ are transitions from (1,1) to (0,1), with probability p_1, and from (1,1) to (1,1) with probability $1-p_1$.

•Events in the process $Z_2(t;\delta_2)$ are transitions from (1,1) to (1,0), with probability p_2, and from (1,1) to(1,1)with probability $1-p_2$.

•Events in the process $Z_{12}(t;\delta_{12})$ are transitions from state (1,1) to state (0,0),(0,1),(1,0),(1,1) with respective probabilities p_{00}, p_{01}, p_{10}, p_{11}.

Let X and Y denote the life length of the first and second components. Since $Z_i(t;\delta_i)$ are independent and have independent increments,

we have for $t \geq s \geq 0$,

$$\begin{aligned} P(X>s,Y>t) &= \left\{\sum_{k=0}^{\infty} e^{\delta_1 s}\frac{(\delta_1 s)^k}{k!}(1-p_1)^k\right\}\left\{\sum_{l=0}^{\infty} e^{\delta_2 t}\frac{(\delta_2 t)^l}{l!}(1-p_2)^l\right\} \\ &\cdot \left\{\sum_{n=0}^{\infty}\sum_{m=0}^{\infty}[e^{\delta_{12}s}\frac{(\delta_{12}s)^m}{m!}p_{11}^m][e^{\delta_{12}(t-s)}\frac{(\delta_{12}(t-s))^n}{n!}(p_{11}+p_{01})^n]\right\} \\ &= \exp\{-s[\delta_1 p_1+\delta_{12}p_{01}]-t[\delta_2 p_2+\delta_{12}(1-p_{11}-p_{01})]\} \quad (2.15) \end{aligned}$$

By symmetry, for $t \geq s \geq 0$,

$$P(X>s,Y>t) = \exp\{-s[\delta_1 p_1+\delta_{12}(1-p_{11}-p_{10})]-t[\delta_2 p_2+\delta_{12}p_{10}]\} \quad (2.16)$$

Consequently, by combining (2.15) and (2.16), it follows that

$$P(X>s,Y>t) = \exp[-\lambda_1 s-\lambda_2 t-\lambda_{12}\max(s,t)], \quad (2.17)$$

where

$$\lambda_1=\delta_1 p_1+\delta_{12}p_{01}, \lambda_2=\delta_2 p_2+\delta_{12}p_{10}, \lambda_{12}=\delta_{12}p_{00}$$

•When $p_1=p_2=1$, $p_{00}=1$, we have the specialized fatal model.

•When $p_1=p_2=0$ we effectively eliminate the first two process; but the joint distribution obtained from the process Z_{12} is of the same form.

2.3.4 Properties of Marshall-Olkin's bivariate exponential distribution

Property 1. Bivariate Memory Less Property

The characterization of the memoryless given in the equation (2.9) is so fundamental in the univariate case, Marshall and Olkin have investigated its bivariate (multivariate) extensions.
If we write $\bar{F}(x, y) = P(X > x, Y > y)$, then a direct extension of (2.9) might be

$$\bar{F}(s_1 + t_1, s_2 + t_2) = \bar{F}(s_1, s_2)\bar{F}(t_1, t_2) \tag{2.18}$$

where $s_1, s_2, t_1, t_2 > 0$
The only solution of the equation (2.18) is:

$$\bar{F}(s, t) = \exp\{(-\theta_1 s - \theta_2 t)\}, \quad \theta_1, \theta_2 > 0 \tag{2.19}$$

Proof
Setting $s_2 = t_2 = 0$ and the, $s_1 = t_1 = 0$, in (2.18) we obtain

$$\bar{F}_1(s_1 + t_1) \equiv \bar{F}(s_)1 + t_1, 0) = \bar{F}_1(s_1)\bar{F}_1(t_1)$$

so that

$$\bar{F}_1(s) = \exp(-\theta_1 s) \text{ with } \quad \theta_1 > 0$$

Similarly

$$\bar{F}_2(t) = \exp(-\theta_2 t) \text{ with } \quad \theta_2 > 0$$

With $\bar{F}_1$ and $\bar{F}_2$ are the survival function of the marginal distributions.
Then taking $t_1 = s_2 = 0$, we obtain

$$\bar{F}(s_1, t_2) = \bar{F}_1(s_1)\bar{F}_2(t_2)$$

from which the results follows.

Property 2. The distribution function

The survival function of BVE given in (2.13) is a mixture of two components, a absolutely continuous part and a singular part (concentrates its mass on the line x=y).

$$\bar{F}(x,y) = \alpha \bar{F}_\alpha(x,y) + (1-\alpha)\bar{F}_s(x,y), \quad 0 \le \alpha \le 1$$

where $\bar{F}_\alpha$ is absolutely continuous and $\bar{F}_s$ is a singular distribution. The weights of the mixture depend on the parameters $\lambda_1, \lambda_2, \lambda_{12}$ and:

$$\alpha = \frac{\lambda_1 + \lambda_2}{\lambda} \quad \text{with} \quad \lambda = \lambda_1 + \lambda_2 + \lambda_{12}$$

where

$$\bar{F}_s(x,y) = \exp[-\lambda \max(x,y)]$$

and

$$\begin{aligned} \bar{F}_\alpha(x,y) &= \frac{\lambda}{\lambda_1 + \lambda_2} \exp[-\lambda_1 x - \lambda_2 y - \lambda_{12} \max(x,y)] \\ &- \frac{\lambda_{12}}{\lambda_1 + \lambda_2} \exp[-\lambda \max(x,y)] \end{aligned}$$

In the form

$$\bar{F}(x,y) = \frac{1}{\lambda}\{(\lambda_1 + \lambda_2)\bar{F}_\alpha(x,y) + \lambda_{12}\bar{F}_s(x,y)\}$$

The density of the absolutely continuous part has the following form:

$$f_\alpha(x,y) = \frac{1}{\alpha}\frac{\partial^2 \bar{F}(x,y)}{\partial x \partial y} = \begin{cases} \lambda_2(\lambda_1 + \lambda_{12})\bar{F}(x,y), & x > y \\ \lambda_1(\lambda_2 + \lambda_{12})\bar{F}(x,y), & x < y \end{cases}$$

Property 3. Probability density function

Bemis et al. (1972) a PDF as follows

$$f(x,y) = \begin{cases} \lambda_1(\lambda_2+\lambda_{12})\exp\{-\lambda_1 x - (\lambda_2+\lambda_{12})y\} & \text{if} \quad 0<x<y \\ \lambda_2(\lambda_1+\lambda_{12})\exp\{-\lambda_2 y - (\lambda_1+\lambda_{12})x\} & \text{if} \quad 0<y<x \\ \lambda_{12}\exp\{-(\lambda_1+\lambda_2+\lambda_{12})y\} & \text{if} \quad 0<y=x \end{cases}$$

Clearly

$$\int_x \int_y f(x,y) = 1$$

Remark

the pdf can be viewed as a mixture of three PDFs with the mixing proportion $\frac{\lambda_1}{\lambda}$, $\frac{\lambda_2}{\lambda}$ and $\frac{\lambda_{12}}{\lambda}$. In fact

$$f_1(x,y) = \begin{cases} \lambda(\lambda_2+\lambda_{12})e^{-\lambda_1 x-(\lambda_2+\lambda_{12})y}, & 0<x<y<\infty \\ 0, & \text{otherwise} \end{cases}$$

$$f_2(x,y) = \begin{cases} \lambda(\lambda_1+\lambda_{12})e^{-(\lambda_1+\lambda_{12})x-\lambda_2 y}, & 0<y<x<\infty \\ 0, & \text{otherwise} \end{cases}$$

and

$$f_3(x,y) = \begin{cases} \lambda e^{-\lambda x}, & 0<y=x<\infty \\ 0, & \text{otherwise} \end{cases}$$

Property 4. Marginal distribution

The marginal distributions are exponentials

$$\bar{F}_1(x) = \bar{F}(x,0) = e^{-(\lambda_1+\lambda_{12})x},$$

$$\bar{F}_2(y) = \bar{F}(0,y) = e^{-(\lambda_2+\lambda_{12})y}$$

Property 5. The conditionals density function

$$f(x|y) = \begin{cases} \lambda_1 e^{-\lambda_1 x}, & x < y \\ \frac{\lambda_2(\lambda_1+\lambda_{12})}{(\lambda_2+\lambda_{12})} e^{-(\lambda_1+\lambda_{12})x+\lambda_{12}x}, & x > y \\ \frac{\lambda_{12}}{(\lambda_2+\lambda_{12})} e^{-\lambda_1 y}, & x = y \end{cases}$$

$$f(y|x) = \begin{cases} \lambda_2 e^{-\lambda_2 y}, & y < x \\ \frac{\lambda_1(\lambda_2+\lambda_{12})}{(\lambda_1+\lambda_{12})} e^{-(\lambda_2+\lambda_{12})y+\lambda_{12}x}, & y > x \\ \frac{\lambda_{12}}{(\lambda_2+\lambda_{12})} e^{-\lambda_2 x}, & y = x \end{cases}$$

• From the survival function and probability density function we can see the the necessary and sufficient condition for X and Y to be independent is that $\lambda_{12} = 0$.

Property 6. Minimum of bivariate exponential random variables

If (X,Y) is distributed as in (2.17) then the survival function of $\min(X, Y)$ has the following form:

$$P(\min(X,Y) > s) = P(X > s, Y > s) = \exp(-(\lambda_1 + \lambda_2 + \lambda_{12})s))$$

Therefore the distribution of $\min(X, Y)$ is exponential as the marginal distributions of X and Y (with different parameter)

Property 7. Representation in terms of independent bivariate exponential variables

Let (X,Y) be distributed according to F. Then F is as in (2.17) if there exist three independent exponential random variables U_1, U_2, U_{12}, such that:

$$X = \min(U_1, U_{12}), \quad Y = \min(U_2, U_{12})$$

Proof

$$\begin{aligned} P(X) > x, Y > y) &= P(U_1 > x, U_{12} > x, U_2 > y, U_{12} > y) \\ &= P(U_1 > x)P(U_2 > y)P(U_{12} > \max(x, y)) \\ &= \exp[-\lambda_1 x - \lambda_2 y - \lambda_{12} \max(x, y)] \end{aligned}$$

By independence of U_1, U_2, U_{12}

Property 8. The survival function of the maximum

The survival function of $\max(X, Y)$ is given by:

$$P(\max(X,Y) > t) = e^{-(\lambda_1+\lambda_{12})t} + e^{-(\lambda_2+\lambda_{12})t} - e^{-(\lambda_1+\lambda_2+\lambda_{12})t}, \quad x > 0$$

Proof

Let T =max(X, Y)

$$\begin{aligned}
P(T \le t) &= P(X \le t, Y \le t) \\
&= \int_0^t \int_0^x \lambda_2(\lambda_1+\lambda_{12}) \exp\{-\lambda_2 y - (\lambda_1+\lambda_{12})x\} dydx \\
&+ \int_0^t \int_x^t \lambda_1(\lambda_2+\lambda_{12}) \exp\{-\lambda_1 x - (\lambda_2+\lambda_{12})y\} dydx + \int_0^t \lambda_{12} \exp\{-\lambda x\} dx \\
&= \int_0^t (\lambda_1+\lambda_{12}) \exp\{-(\lambda_1+\lambda_{12})x\}(1-\exp\{-\lambda_2 x\})dx \\
&+ \int_0^t \lambda_1 \exp\{-\lambda_1 x\} \left\{\exp\{-(\lambda_2+\lambda_{12})x\} - \exp\{-(\lambda_2+\lambda_{12})t\}\right\} dx + \frac{\lambda_{12}}{\lambda}(1-\exp\{-\lambda t\}) \\
&= 1 - \exp\{-(\lambda_1+\lambda_{12})t\} - \frac{(\lambda_1+\lambda_{12})}{\lambda}(1-\exp\{-\lambda t\}) \\
&+ \frac{\lambda_1}{\lambda}(1-\exp\{-\lambda t\}) \exp\{-(\lambda_1+\lambda_{12})t\}(1-\exp\{-\lambda_1 t\} \\
&+ \frac{\lambda_{12}}{\lambda}(1-\exp\{-\lambda t\}) \\
&= 1 - \exp\{-(\lambda_1+\lambda_{12})t\} + \exp\{-(\lambda_2+\lambda_{12})t\} - \exp\{-(\lambda_1+\lambda_2+\lambda_{12})t\}
\end{aligned}$$

Property 9.

Let (X_i, Y_i), i=1,2,...,n be random sample from BVE distribution with parameters $(\lambda_1, \lambda_2, \lambda_{12})$.
IF
N_1: number of pairs (X_i, Y_i) such that $X_i < Y_i$,
N_2: number of pairs (X_i, Y_i) such that $X_i > Y_i$,
N_{12}: number of pairs (X_i, Y_i) such that $X_i = Y_i$,

Then

$$(N_1, N_2) \sim Trinomial\left(N, \frac{\lambda_1}{\lambda}, \frac{\lambda_2}{\lambda}\right)$$

where $N = N_1 + N_2 + N_{12}$.

Property 10. Moment generating function of MOBVE

• The moment generating function of BVE is given by

$$\psi(s,t) = \int_0^\infty \int_0^\infty e^{-sx-ty} dF(x,y) = \frac{(\lambda+s+t)(\lambda_1+\lambda_{12})(\lambda_2+\lambda_{12})(st\lambda_{12})}{(\lambda+s+t)(\lambda_1+\lambda_{12}+s)(\lambda_2+\lambda_{12}+t)}$$

where $\lambda = \lambda_1 + \lambda_2 + \lambda_{12}$

•The moment of MOBVE are:

$$E(X) = \frac{1}{\lambda_1+\lambda_{12}}, \quad V(X) = \frac{1}{(\lambda_1+\lambda_{12})^2}$$

$$E(Y) = \frac{1}{\lambda_2+\lambda_{12}}, \quad V(Y) = \frac{1}{(\lambda_2+\lambda_{12})^2}$$

$$E(XY) = \frac{1}{\lambda}(\frac{1}{\lambda_1+\lambda_{12}} + \frac{1}{\lambda_2+\lambda_{12}})$$

Hence the covariance is given by

$$Cov(X,Y) = \frac{\lambda_{12}}{\lambda(\lambda_1+\lambda_{12})(\lambda_2+\lambda_{12})}$$

The correlation coefficient is:

$$\rho(X,Y) = \frac{\lambda_{12}}{\lambda}$$

We can note that

$$\begin{aligned} \rho(X,Y) &= P(X=Y) = P\{first\ fatal\ shocks\ affects\ both\ components\} \\ &= \frac{\lambda_{12}}{\lambda} = \frac{\lambda_{12}}{\lambda_1+\lambda_2+\lambda_{12}} > 0 \end{aligned}$$

•The distribution of (X,Y) is not absolutely continuous with respect to the usual Lebesgue measure since $\rho(X,Y) = P(X = Y)$

2.3.5 Comparison of MOBVE with independence case

Often in reliability theory, it is assumed that the components of a system have independent life lengths. Marshall and Olkin have studied the effect of this assumption in their distribution BVE.
The marginal distributions are given by:

$$\bar{F}_1(x) = e^{-(\lambda_1+\lambda_{12})x}, \quad \bar{F}_2(y) = e^{-(\lambda_2+\lambda_{12})y}$$

They suppose that the operate under the assumption the joint distribution F(x,y) is $F_1(x)F_2(y)$, when in fact $\bar{F}(x,y)$ is given by (3.1). The difference

$$\bar{F}(x,y) - \bar{F}_1(x)\bar{F}_2(y) = e^{-\lambda_1 x - \lambda_2 y - \lambda_{12}\max(x,y)}(1 - e^{-\lambda_{12}\min(x,y)})$$

is positive for all x and y.
So the probability that both items survive is actually greater than the assumption of independence would lead us to believe.
However, for any bivariate distribution F with marginals F_1 and F_2

$$\bar{F}(x,y) - \bar{F}_1(x)\bar{F}_2(y) = F(x,y) - F_1(x)F_2(y)$$

So the probability that both items fail is also greater than that computed under the assumption of independence.
This means that:
• in the case of series system (which functions only when both items function), is greatest in the case of dependence.
• in the case of a parallel system (which fails only when both items fail), is greatest in the case of independence.

2.3.6 Transformation of Marginals

Transformation to Weibull Marginals

Some distributions can be obtained by a change of variables, for example:
If (X,Y) are BVE , the distribution of $(X^{1/\beta}, Y^{1/\gamma})$ is:

$$\bar{F}(x, y) = \exp[-\lambda_1 x^\beta - \lambda_2 y^\gamma - \lambda_{12} \max(x^\beta, y^\gamma)]$$

which is the Marshall-Olkin bivariate Weibull distribution denoted MOBW where the marginals are Weibull distribution.

Approach of Muliere and Scarsini

First, consider the univariate case. Muliere and Scarsini (1987) presented a general version of the lack of memory property as follows:

$$\bar{F}(s * t) = \bar{F}(s)\bar{F}(t) \tag{2.20}$$

where $*$ is any binary operation that is associative (i.e., such that (x∗y)∗z =x∗(y∗z)). Examples include the following:

•The operation $*$ being addition leads to the usual lack of memory characterization of exponential distribution given in (2.18).

•If x∗y =$(x^\alpha + y^\alpha)^{1/\alpha}$, then the Weibull distribution $\bar{F}(x) = exp(-\lambda x^\alpha)$ results.

•If x∗y = xy, then the Pareto distribution $\bar{F}(x) = x^{-\lambda}$ results.

In the bivariate case, consider first the following version of the bivariate lack of memory property:

$$\bar{F}(s_1 + t, s_2 + t) = \bar{F}(s_1, s_2)\bar{F}(t, t)$$

Now consider:

$$\bar{F}(s_1 * t, s_2 * t) = \bar{F}(s_1, s_2)\bar{F}(t, t), \tag{2.21}$$

The solution of (2.20) is given by:

$$\bar{F}(s, t) = \exp(-\lambda_1 s - \lambda_2 t - \lambda_{12} g[\max(s, t)]) \tag{2.22}$$

with g(.) being a (strictly increasing) function corresponding to the operation $*$; i.e., g(x$*$y) = g(x)+g(y).
Examples include the following:
• The operation $*$ being addition leads to the Marshall and Olkin distribution.
• If x$*$y = $(x^\alpha + y^\alpha)^{1/\alpha}$, the Weibull version of the Marshall and Olkin distribution results, i.e., $\bar{F}(x,y) = exp[-\lambda_1 x^\alpha - \lambda_2 y^\alpha - \lambda_{12}\max(x^\alpha, y^\alpha)]$.
•If x$*$y = xy, then the result is $\bar{F}(x,y) = x^{-\lambda_1} y^{\lambda_2}[max(x,y)]^{\lambda_{12}}$, the Pareto version of Marshall and Olkin's distribution.

2.3.7 Some studies on MOBVE model

• Block (1979), proved that X and Y with exponential marginals have Marshall and Olkin's bivariate distribution if and only if one of the following two equivalent conditions holds:

1. $\min(X,Y)$ has an exponential distribution,
2. X-Y and $\min(X,Y)$ are independent.

•Nadarajah and Kotz (2005) studied the exact distribution of R=X+Y and W=X/(X+Y) when X and Y are BVE. They showed that the pdfs of R and w are given by

$$\begin{aligned} f_R(r) &= \frac{\lambda_1(\lambda_2+\lambda_{12})}{\lambda_{12}+\lambda_2-\lambda_1}\exp\{-(\lambda_2+\lambda_{12})r\}[\exp\{(1/2)(\lambda_2+\lambda_{12}-\lambda_1)r\}-1] \\ &+ \frac{\lambda_1(\lambda_2+\lambda_{12})}{\lambda_{12}+\lambda_2-\lambda_1}\exp\{-(\lambda_2+\lambda_{12})r\}[1-\exp\{-(1/2)(\lambda_2-\lambda_{12}-\lambda_1)r\}] \\ &+ \lambda_{12}\exp\{-(1/2)(\lambda_1+\lambda_2+\lambda_{12})r\} \end{aligned}$$

and

$$f_W(w) = \begin{cases} \frac{\lambda_1(\lambda_2+\lambda_{12})}{\{\lambda_1 w+(\lambda_2+\lambda_{12})(1-w)\}^2}, & \text{if} \quad w < 1/2 \\ \frac{\lambda_2(\lambda_1+\lambda_{12})}{\{\lambda_2(1-w)+(\lambda_1+\lambda_{12})w\}^2}, & \text{if} \quad w > 1/2 \\ \frac{\lambda_{12}}{(\lambda_1+\lambda_2+\lambda_{12})^2(1-w)^2}, & \text{if} \quad w = 1/2 \end{cases}$$

respectively, for $0 < r < \infty$, and $0 < w < 1$. They are also derived the corresponding moments.
•Ryu (1993) extended Marshall and Olkin's model such that the new joint distribution is absolutely continuous and need not be memoryless. The survival function of this extension

is given by:

$$\begin{aligned}\bar{F}(x,y) &= \exp[-(\lambda_1+\lambda_{12})x-\lambda_2 y+\frac{\lambda_{12}}{s_1}\{1-\exp(-s_1x+s_1y)\}\\ &+ \frac{\lambda_{12}}{s_1+s_2}\{\exp(-s_1x+s_1y)-\exp(-s_1x-s_2y)\}]\end{aligned}$$

for x>y, and

$$\begin{aligned}\bar{F}(x,y) &= \exp[-\lambda_1x-(\lambda_2+\lambda_{12})y+\frac{\lambda_{12}}{s_2}\{1-\exp(-s_2y+s_2x)\}\\ &+ \frac{\lambda_{12}}{s_1+s_2}\{\exp(-s_2y+s_2x)-\exp(-s_1x-s_2y)\}] \; for\ x<y.\end{aligned}$$

• The exact distribution of the product XY is given in Nadarajah (2006).
• An expression for Rényi and Shannon entropy for this distribution was obtained by Nadarajah and Zografos (2005).

2.3.8 Application fields

Among many applications of Marshall and Olkin's bivariate distribution, we note especially the nuclear reactor safety, competing risks and reliability. Certainly, the idea of simultaneous failure of two components is far from being merely of academic interest.
•Hagen (1980) has presented a review in the context of nuclear power and has pointed out that introducing redundancy into a system reduces random component failure to insignificance, leading to the common-mode/common-cause type being predominant among system failure.

Some examples of application in reliability:

• Ramanarayanan and Subramanian (1981) reported on a two-component system with Marshall and Olkin distributions of both life and repair times.
• Jeevanand (1998) used Marshall and Olkin distributions to describe Bayes estimation of reliability under the stress strength model.
• Klefsjo (1982) studied certain engineering and biological systems consisting of two components, each with its own lifetime. The joint probability distribution of lifetimes was modeled by the Marshall and Olkin's distribution.
•Osaki et al. (1989) provided some results for availability measures of systems in which two units are in series, failure of unit 1 shuts off unit 2, but not vice versa, the lifetimes

follow Marshall and Olkin's distribution, the units have arbitrary repair-time distributions, and two alternative assumptions are made about the repair discipline.
•Among other areas of application, Rai and Van Ryzin (1984) applied Marshall and Olkin's distribution as a tolerance distribution in a quantal response context, to the occurrence of bladder and liver tumors in mice exposed to one of several alternative dosages of a carcinogen. Actually, Marshall and Olkin's distribution was (i) used in the form with Weibull marginals, and (ii) mixed with a finite probability of tumors occurring even at zero dose.

2.3.9 Comparison of the Marshall-Olkin distributions with Freund and Block and Basu distributions

	Marshall and Olkin	Block and Basu	Freund
Exponential marginals	+		
Absolutely continuous		+	+
Bivariate lack of memory	+	+	+
min(X,Y) is exponential	+	+	+
min(X,Y) is independent of X-Y	+	+	+

Table 2.1: Comparison of the Marshall-Olkin distributions with Freund and Block and Basu distributions

Chapter 3

Inference for Marshall-Olkin model

Among several bivariate (multivariate) exponential distribution, the Marshall-Olkin bivariate distribution is the only bivariate distribution with exponential marginals and bivariate lack of memory property. Statistical inferences about the Marshall-Olkin distribution has been discussed in the literature. Arnold (1968) has obtained unbiased estimators which are uniformly minimum variance unbiased estimators when components are in series. Bemis et al. (1972) have presented maximum likelihood estimators (MLEs) and tests for the correlation between X and Y. Proschan and Sullo (1974) have also presented MLEs of $(\lambda_1, \lambda_2, \lambda_{12})$ for the series and parallel systems. Unfortunately, statistical inference for MOBVE is not easy task due the complicated form of its density function because the MOBVE is not absolutely continuous. Another disadvantage of this model is that in practice it is difficult to obtain large samples, situations with small samples may presented difficult results to justify. We explore the use of Bayesian method as an alternative to classical analysis. When we get new information, we should update our probabilities to take the new information into account. Bayesian methods tell us exactly how to do this. We start this chapter with the preliminary notions in Bayesian paradigm.

3.1 Preliminary notions in Bayesian paradigm

3.1.1 Bayes theorem

If A and E are events such that $P(E) \neq 0$, P(A|E) and P(E|A) are related by

$$P(A|E) = \frac{P(E|A)P(A)}{P(E)}$$

Thomas Bayes (1764) has proven a continuous version of this result, namely, that given two random variables x and y, with conditional distribution f(x|y) and marginal distribution g(y), the conditional distribution of y given x is

$$g(y|x) = \frac{f(x|y)g(y)}{\int f(x|y)g(y)dy}$$

3.1.2 Definition

A parametric Bayesian statistical model is made of a statistical model which consists of the observation of a random variable x, distributed according to $f(x|\theta)$, and a prior distribution on the parameters, $\pi(\theta)$.
Given these two distributions, we can construct several distributions, namely
(a) ***the joint distribution of (θ,x),***

$$f(x, \theta) = f(x|\theta)\pi(\theta)$$

(b)***the marginal distribution of x***

$$f(x) = \int_{\Theta} f(x|\theta)\pi(\theta)d\theta$$

(c)***Posterior distribution,*** the inference is based on the distribution of θ conditional on x, $\pi(\theta|x)$, called posterior distribution which is obtained by using the continuous version of Bayes's formula

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{\int_{\Theta} f(x|\theta)\pi(\theta)d\theta}$$

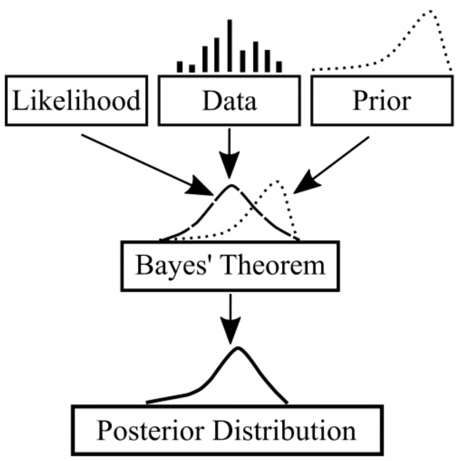

Figure 3.1: The Bayesian procedure

Let's define the spaces intervening in a Bayesian model:

• **Observations space:** denoted X, it represents all the results following a study of a phenomenon.

• **Actions space:** denoted A, it represents all the actions or decisions to be taken after obtaining the information.

• **States of nature space:** denoted Θ, it represents the space of the unknown parameters θ.

• **Decision rules space:** denoted D, it represents the set of decision rules that we define as an application from X to A denotes by δ:

$$\delta : X \rightarrow A, \quad x_i \rightarrow \delta(x_i) = a_i$$

a_i being th i^{th} action.

3.1.3 Loss function and risk

Loss function

A loss function is any function L from $(\Theta \times D)$ in $[0, +\infty[$

$$L(\Theta, D) \to \mathbb{R}^+$$

$$(\theta, \delta) \to L(\Theta, \delta(x))$$

This loss function is supposed to evaluate the penalty (or error) $L(\theta, \delta)$ associated with the decision d when the parameter takes the value θ.
The actual determination of the loss function is often awkward in practice, in particular because the determination of the consequences of each action for each value of θ is usually impossible when D or Θ are large sets, for instance when they have an infinite number of elements.

Frequentist risk

Definition In order to derive an effective comparison criterion from the loss function, the frequentist approach proposes to consider instead the average loss or frequentist risk denoted $R(\theta, \delta(x))$ given by

$$R(\theta, \delta(x)) = E_\theta[L(\theta, \delta(x))] = \int_X L(\theta, \delta(x)) f(x|\theta) dx$$

Posterior risk

Definition On the contrary, the Bayesian approach to Decision Theory integrates on the space Θ since θ is unknown, instead of integrating on the space X as x is known. It relies on the posterior expected loss $\rho(\pi, \delta(x))$

$$\rho(\pi, \delta(x)) = E^{\pi(.|x)}[L(\theta, \delta(x))] = \int_\Theta L(\theta, \delta(x)) \pi(\theta|x) d\theta$$

Integrated risk

Definition Given a prior distribution π, it is also possible to define the integrated risk $r(\pi, \delta(x)$, which is the frequentist risk averaged over the values of θ according to their prior distribution

$$r(\pi, \delta(x)) = E^{\pi}[R(\theta, \delta(x))] = \int_{\Theta} \int_{X} L(\theta, \delta(x)) f(x|\theta) \pi(\theta) dx d\theta.$$

Bayes risk

Definition A Bayes estimator associated with a prior distribution π and a loss function L is any estimator δ^{π} which minimizes $r(\pi, \delta(x))$ For every $x \in X$, it is given by $\delta^{\pi}(x)$, argument of $\min_d \rho(\pi, \delta(x))$. The value $r(\pi) = r(\pi, \delta^{\pi}(x))$ is then called the Bayes risk.
•A fundamental basis of Bayesian Decision Theory is that statistical inference should start with the rigorous determination of three factors:

1. the distribution family for the observations,$f(x|\theta)$;
2. the prior distribution for the parameters, $\pi(\theta)$;
3. the loss associated with the decisions, $L(\theta, \delta(x))$

3.1.4 Usual loss functions

The quadratic loss

Definition The quadratic loss function is defined by:

$$L(\theta, \delta(x)) = (\theta - \delta(x))^2$$

Proposition The Bayes estimator δ^{π} associated with the prior distribution π and with the quadratic loss is the posterior expectation

$$\delta^{\pi}(x) = E^{\pi}[\theta|x]$$

The following corollaries are straightforward to derive

Corollary 1 The Bayes estimator δ^π associated with π and with the weighted quadratic loss

$$L(\theta, \delta(x)) = \omega(\theta)(\theta - \delta(x))^2$$

where $\omega(\theta)$ is a nonnegative function, is

$$\delta^\pi(x) = \frac{E^\pi[\omega(\theta)\theta|x]}{E^\pi[\omega(\theta)|x]}$$

Corollary 2 When $\Theta \in \mathbb{R}^p$ the Bayes estimator δ^π associated with π and with the quadratic loss,

$$L(\theta, \delta(x)) = (\theta - \delta(x))^t Q(\theta - \delta(x))$$

is the posterior mean

$$\delta^\pi(x) = E^\pi[\theta|x],$$

for every positive-definite symmetric $p \times p$ matrix Q.

The absolute error loss

Definition An alternative solution to the quadratic loss in dimension one is to use the absolute error loss,

$$L(\theta, \delta(x)) = |\theta - \delta(x)|$$

Proposition The Bayes estimator δ^π associated with the prior distribution π and with the absolute error loss is the posterior median.
More generally, a multilinear function

$$L(\theta, \delta(x)) = \begin{cases} k_2(\theta - \delta(x)); & \theta > \delta(x) \\ k_1(\delta(x) - \theta); & \theta < \delta(x) \end{cases}$$

Proposition A Bayes estimator associated with the prior distribution π and the multilinear loss is a $(k_2/(k_1 + k_2))$ fractile of $\pi(\theta|x)$.

The 0 - 1 loss

Definition This loss is mainly used in the classical approach to hypothesis testing. More generally, this is a typical example of a non quantitative loss. The 0 - 1 loss is defined by

$$L(\theta, \delta(x)) = \begin{cases} 0; & (\theta - \delta(x)) < \epsilon \\ 1; & (\theta - \delta(x)) > \epsilon \end{cases}$$

Proposition A Bayes estimator associated with the prior distribution π and the 0 - 1 loss is the mode of $\pi(\theta|x)$.

3.1.5 The choice of prior distributions

Undoubtedly, the most critical and most criticized point of Bayesian analysis deals with the choice of the prior distribution, since, once this prior distribution is known, inference can be led in an almost mechanic way by minimizing posterior losses. The prior distribution is the key to Bayesian inference and its determination is therefore the most important step in drawing this inference. To some extent, it is also the most difficult. Indeed, in practice, it seldom occurs that the available prior information is precise enough to lead to an exact determination of the prior distribution, in the sense that many probability distributions are compatible with this information.Most often, it is then necessary to make a (partly) arbitrary choice of the prior distribution. We will consider two common techniques, the conjugate prior approach which requires a limited amount of information, and the noninformative approach, which can be directly derived from the sampling distribution.

The conjugate prior approach

Definition A family F of probability distributions on Θ is said to be conjugate (or closed under sampling) for a likelihood function $f(x|\theta)$ if, for every $\pi \in F$, the posterior distribution $\pi(\theta|x)$ also belongs to F.
Conjugate prior distributions are usually associated with a particular type of sampling distribution that always allows for their derivation, and is even characteristic of conjugate priors, as we will see below. These distributions constitute what is called exponential families.

Definition A family of distributions is called an exponential family with s parameters if their density function have the following form

$$f(x|\theta) = \exp[\Sigma_{i=1}^{s}\eta_i(\theta)T_i(\theta) - B(\theta)]h(x)$$

with $\eta_i(.)$ et B(.) are functions with parameter θ et $T_i(.)$ are statistics.
Proposition A conjugate family for $f(x|\theta)$is given by

$$\pi(\theta|\mu,\lambda) = K(\mu,\lambda)\exp(\theta\mu - \lambda A(\theta))$$

with $K(\mu,\lambda)$ is the normalizing constant of the density.
The corresponding posterior distribution is

$$\pi(\theta|\mu + x, \lambda + 1)$$

f(x|θ)	$\pi(\theta)$	$\pi(\theta\|x)$
N($\theta\|\sigma^2$)	N(μ,τ^2)	N($x/\sigma^2 + \mu/\tau^2, [1/\sigma^2 + 1/\tau^2]^{-1}$)
G(n,θ)	G(α,β)	G($\alpha + n, \beta + x$)
B(n,θ)	Beta(α,β)	Beta($\alpha + n, \beta + x$)
P(θ)	G(α,β)	G($\alpha + x, \beta + 1$)

Table 3.1: Examples of conjugate distribution

The noninformative prior approach

This approach is used when no prior information is available.

•**Improper prior distributions:** the prior distribution is said to be improper (or generalized) if:

$$\int_{\Theta}\pi(\theta)d\theta = \infty$$

or

$$\sum_{\theta\in\Theta}\pi(\theta) = \infty$$

•**Laplace's prior:** Historically, Laplace was the first to use noninformative techniques, since, although he had no information about the number of white balls in the urn or the proportion of male births, he put a prior distribution on these parameters that took into account his ignorance and gave the same likelihood to each value of the parameter, i.e., he used a uniform prior. His reasoning, later called the Principle of Insufficient Reason, was based on the equiprobability of elementary events and therefore appeared to be sound enough.

•**The Jeffreys prior:** This method was described by Jeffreys (1946) and it is based on the Fisher information given by

$$I(\theta) = E[-\frac{\partial^2}{\partial\theta^2}\log f(x|\theta)]$$

Jeffreys prior is defined as

$$\pi(\theta) = C\sqrt{I(\theta)} \quad (\pi(\theta) \propto \sqrt{I(\theta)})$$

$\star$ When $\theta \in \Theta_q$ the Jeffreys prior is given by

$$\pi(\theta) = [detI(\theta)]^{1/2}$$

with $I(\theta)$is the Fisher matrix:

$$I(\theta)|_{i,j} = E\left[-\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log f(x|\theta)\right], 1 \leq i,j \leq q$$

3.1.6 Credible intervals

As in the testing setting, the Bayesian paradigm proposes a notion of confidence regions that is more natural that its frequentist counterpart since, again, the notation $P(\theta \in C_x)$

is meaningful even conditional upon x.
Definition For a prior distribution π, a set C_x is said to be an α-credible set if

$$P^{\pi}(\theta \in C_x|x) \geq 1 - \alpha$$

This region is called an HPD α-credible region (highest posterior density) if it can be written under the form

$$\{\theta; \pi(\theta|x) > k_{\alpha}\} \subset C_x^{\pi} \subset \{\theta; \pi(\theta|x) \geq k_{\alpha}\}$$

where k_{α} is the largest bound such that

$$P^{\pi}(\theta \in C_x^{\alpha}|x) \geq 1 - \alpha$$

3.1.7 MCMC methods

One of the simplest and most powerful practical uses of the ergodic theory of Markov chains is in Markov chain Monte Carlo (MCMC). Suppose we wish to simulate from a probability density π (which is called the target density) but that direct simulation is either impossible or practically infeasible (possibly due to the high dimensionality of π). This generic problem occurs in diverse scientific applications, for instance Statistics, Computer Science, and Statistical Physics.
Markov chain Monte Carlo offers an indirect solution based on the observation that it is much easier to construct an ergodic Markov chain with π as a stationary probability measure, than to simulate directly from π.
Suppose that π is a (possibly unnormalised) density function, with respect to some reference measure (e.g. Lebesgue measure, or counting measure) on some state space X . Assume that π is so complicated, and X is so large, that direct numerical integration is infeasible. We now describe several MCMC algorithms, which allow us to approximately sample from π. In each case, the idea is to construct a Markov chain update to generate X_{t+1} given X_t, such that π is a stationary distribution for the chain, i.e. if X_t has density π, then so will X_{t+1}.

The Metropolis-Hastings algorithm

The Metropolis-Hastings algorithm proceeds in the following way. An initial value X_0 is chosen for the algorithm. Given X_t , a candidate transition Y_{t+1} is generated according to some fixed density $q(X_t, .)$, and is then accepted with probability $\alpha(X_t, Y_{t+1})$, given by

$$\alpha(x, y) = \begin{cases} \min\left\{\frac{\pi(y)q(y,x)}{\pi(x)q(x,y)}, 1\right\}; & \pi(x)q(x, y) > 0 \\ 1; & \pi(x)q(x, y) = 0 \end{cases}$$

otherwise it is rejected. If Y_{t+1} is accepted, then we set $X_{t+1} = Y_{t+1}$. If Y_{t+1} is rejected, we set $X_{t+1} = X_t$. By iterating this procedure, we obtain a Markov chain realization $X = \{X_0, X_1, X_2, ...\}$.

The Metropolis-Hastings algorithm was chosen precisely to ensure that, if X_t has density π, then so does X_{t+1}. Thus, π is stationary for this Markov chain. It then follows from the ergodic theory of Markov chains that, under mild conditions, for large t, the distribution of X_t will be approximately that having density π. Thus, for large t, we may regard X_t as a sample observation from π.

Note that this algorithm requires only that we can simulate from the density $q(x, .)$ (which can be chosen essentially arbitrarily), and that we can compute the probabilities $\alpha(x, y)$. Further note that this algorithm only ever requires the use of ratios of π values, which is convenient for application areas where densities are usually known only up to a normalization constant, including Bayesian Statistics and Statistical Physics.

The Gibbs sampler

Assume that $X = R^d$, and that π is a density function with respect to d-dimensional Lebesgue measure. We shall write $x = (x^{(1)}, x^{(2)}, x^{(3)}, ...x^{(d)})$ for an element of R^d where $x^{(i)} \in R$, for $1 \leq i \leq d$.We shall also write $x^{(-i)}$ for any vector produced by omitting the ith component, $x^{(-i)} = (x^{(1)}, ..., x^{(i-1)}, x^{(i+1)}, .., x^{(d)})$, from the vector x.

The idea behind the Gibbs sampler is that even though direct simulation from p may not be possible, the 1-dimensional conditional densities $\pi_i(.|x^{(-i)})$, for $1 \leq i \leq d$, may be much more amenable to simulation. This is a very common situation in many simulation examples, such as those arising from the Bayesian analysis of hierarchical models.

The Gibbs sampler proceeds as follows. Let P_i be the Markov chain update that, given $X_t = x$, samples $X_{t+1} \sim \pi_i(.|x^{(-i)})$, as above. Then the systematic-scan Gibbs sampler is given by $P = P_1 P_2 ... P_d$, while the random-scan Gibbs sampler is given by $P = (P_1 + P_2 + ... + P_d)/d$.

Importance sampling

Importance sampling (IS) refers to a collection of Monte Carlo methods where a mathematical expectation with respect to a target distribution is approximated by a weighted average of random draws from another distribution. Together with Markov Chain Monte Carlo methods, IS has provided a foundation for simulation-based approaches to numerical integration since its introduction as a variance reduction technique in statistical physics. The appeal of IS lies in a simple probability result. Let p(x) be a probability density for a random variable X and suppose we wish to compute an expecation $I_f = E_p[f(x)]$, with

$$I_f = \int f(x)p(x)dx$$

Then for any probability density $q(x)$ that satisfies $q(x) > 0$ whenever $f(x)p(x) \neq 0$, one has

$$I_f = E_q[w(x)f(X)]$$

where $w(x) = \frac{p(x)}{q(x)}$ and now $E_q[.]$ denotes the expectation with respect to $q(x)$.
Therefore a sample of independent draws $x^{(1)}, ..., x^{(m)}$ from $q(x)$ can be used to estimate I_f by

$$I_f = \frac{1}{m} \sum_{j=1}^{m} w(x^j) f(x^j).$$

In many applications the density $p(x)$ is known only up to a normalizing constant. Here one has $w(x) = cw_0(x)$ where $w_0(x)$ can be computed exactly but the multiplicative constant c is unknown. In this case one replaces I_f with the ratio estimate

$$I_f = \frac{\sum_{j=1}^{m} w(x^j) f(x^j)}{\sum_{j=1}^{m} w(x^j)}$$

3.2 Classical inference for the Marshall-Olkin Model

The Marshall and Olkin (1967) bivariate exponential distribution (MOBVE) has a joint survival function

$$\bar{F}(x,y) = \exp[-\lambda_1 x - \lambda_2 y - \lambda_{12}\max(x,y)] \quad x,y>0 \tag{3.1}$$

For this model the marginal distributions of (X,Y) are exponential with hazard rates $(\lambda_1+\lambda_{12})$ and $(\lambda_1+\lambda_{12})$

To estimate the parameters of the MOBVE suppose we have a sample of size N of (3.1)

$*$Let N_1, N_2 and N_{12} the number of observations with {X<Y},{X>Y} , and {X = Y}, respectively.

and

$$S_1 = \sum_j X_j, \quad S_2 = \sum_j Y_j \quad T = \sum_j \min(X_j,Y_j), \quad T' = \sum_j \max(X_j,Y_j)$$

$*$ We note that (N_1, N_1, N_{12}) have a Multinomial $(n, \frac{\lambda_1}{\lambda}, \frac{\lambda_2}{\lambda}, \frac{\lambda_{12}}{\lambda})$ and S_1, S_2 and T' have gamma distribution with scale parameter $\lambda_1+\lambda_{12}$, $\lambda_1+\lambda_{12}$, and λ and common shape parameter n.

• Using the underlying structure of MOBVE and their properties, Arnold (1968) has suggested a reasonably simple procedure to estimate the hyper parameter of this bivariate distribution. The fact that the maximum likelihood estimation is not easy because the MOBVE distribution is singular and the method of moments become difficult to even write down.

Arnold (1968) has considered estimation when only the minimum of X and Y is observed. He shows that in this case N_1, N_2 and T are complete sufficient statistics and the minimum variance unbiased estimators of $(\lambda_1, \lambda_2, \lambda_{12})$ are

$$\boxed{\widehat{\lambda}_i = \frac{n-1}{n}\frac{N_i}{T} \quad i = 1,2,12}$$

They are consistent estimators.

$*$ The maximum likelihood estimation is considered by several authors including Bemis et al(1972), Proschan and Sullo (1974, 1976), and Bhattacharyya and Johnson (1971, 1973).

In this case the log-likelihood function:

$$\begin{aligned}\log L &= -\lambda_1 S_1 - \lambda_2 S_2 - \lambda_{12} T' + N_1 \log\{\lambda_1(\lambda_1 + \lambda_{12})\} \\ &+ N_2 \log\{\lambda_2(\lambda_1 + \lambda_{12})\} + N_{12} \log \lambda_{12}\end{aligned}$$

Maximum likelihood estimates (MLEs) are found by solving the score equations, namely

$$N_1/\lambda_1 + N_2/\lambda_1 + \lambda_{12} - S_1 = 0,$$

$$N_2/\lambda_2 + N_1/\lambda_2 + \lambda_{12} - S_2 = 0,$$

and

$$N_1/\lambda_2 + \lambda_{12} + N_2/\lambda_1 + \lambda_{12} + N_{12}/\lambda_{12} - T' = 0$$

*If all the $N_i > 0$, then the maximum likelihood estimators of the$\lambda's$ are the unique solution to these equations which is found numerically.
*If $N_{12} = 0$ then $(\widehat{\lambda}_1 \widehat{\lambda}_2, \widehat{\lambda}_{12}) = (\frac{n}{S_1}, \frac{n}{S_2}, 0)$ provided that $N_1 N_2 > 0$
*If $N_{12} = 0$ and either $N_1 = 0$ or $N_2 = 0$the MLE exists but is not unique
*If $N_{12} > 0$ and either $N_1 = 0$ or $N_2 = 0$ the MLE does not exist.
• Proschan and Sullo (1976) propose an estimator based on the first iteration in a maximization of the likelihood equation. Their estimators are

$$\boxed{\widehat{\lambda'}_i = \frac{nN_i}{(N_1 + N_2)S_i} \quad (i = 1, 2)}$$

and

$$\boxed{\widehat{\lambda'}_{12} = \frac{n}{T'}\left\{1 + \frac{N_1}{N_1 + N_2} + \frac{N_2}{N_1 + N_2}\right\}}$$

• Bemis, Bain, and Higgins (1972) give the method of moments estimators, namely,

$$\boxed{\widetilde{\lambda}_i = \frac{\frac{n}{S_i} - \frac{N_{12}}{S_{3-i}}}{1 + \frac{N_{12}}{n}} \quad i = (1,2)}$$

and

$$\boxed{\widetilde{\lambda}_{12} = \frac{N_{12}(\frac{1}{S_1} + \frac{1}{S_2})}{1 + \frac{N_{12}}{n}}}$$

Remark The asymptotic joint distribution of each these sets of estimators are trivariate normal.

• A special case of the MOBVE distribution is the symmetric case ie $\lambda_1 = \lambda_2 = \theta(say)$. Bhattacharyya and Johnson (1971) showed that, in this case, the MLE's of θ and λ_{12} are uniquely determined if $N_{12} < n$ and are given by

$$\widehat{\theta} = \begin{cases} \frac{2n}{S_1+S_2} & if\ N_{12} = 0 \\ \frac{(n-N_{12})\widehat{\lambda}_{12}}{N_{12}+\widehat{\lambda}_{12}S_1} & if\ N_{12} > 0 \end{cases}$$

and

$$\widehat{\lambda}_{12} = \begin{cases} 0 & if\ N_{12} = 0 \\ \frac{\{n^2(S_1-S_2)^2+4N_{12}S_{11}S_2\}^{1/2}+n(S_1-S_2}{2S_1S_2} & if\ N_{12} > 0 \end{cases}$$

3.3 Tests of independence

A variety of authors have considered the inference problems based on the Marshall-Olkin Bivariate exponential model. In this section we shall consider test for the hypothesis of independence, $\lambda_{12} = 0$.
the problem of testing the hypothesis

$$H_0 : \lambda_{12} = 0 \Rightarrow \rho = \frac{\lambda_{12}}{\lambda} = 0$$

against

$$H_1 : \lambda_{12} > 0$$

• Basu (1981), if we observe any observations along the diagonal where X = Y then we must have $\lambda_{12} > 0$, and the null hypothesis should be rejected. The probability of seeing a pair with $X_i = Y_i$ is ρ. This test has a Type I error of zero and power of $1-(1-\rho)^n$.
• Bemis, et al (1972) show that the most powerful test of H_0 against
$H_1 : \rho = \rho_0 > 0$, when λ_1 and λ_2 are known, rejects if either $N_{12} > 0$ or if

$$T = \frac{\rho_0}{1+\rho_0}\left[(\delta_1+\delta_2)S_1 + N_1 \ln\left(\frac{1-\rho_0\delta_2/\delta_1}{1-\rho_0\delta_1/\delta_2}\right)\right] > C(\alpha)$$

where $\delta_i = \lambda_i + \lambda_{12}$, i=1,2.
The critical value $C(\alpha)$ is determined from the power function of the test, namely,

$$Q(\rho) = 1-(1-\rho)^n + (1-\rho)^n \sum_{i=1}^{n} P[N_1 = j|N_{12} = 0]P[T > C(\alpha)|N_1 = j, N_{12} = 0]$$

Here

$$P[N_1 = j|N_{12} = 0] = \binom{n}{j}(\delta_1 - \rho\delta_2)^j(\delta_2 - \rho\delta_1)^{n-j}[(\delta_1+\delta_2)(1-\rho)]^{-n}$$

and

$$P[T > C(\alpha)|N_1 = j, N_{12} = 0] = P\left[\chi_2^{2n} > 2(1+\rho^{-1})(1+\rho)\rho^{-1}\left\{C(\alpha) - j\ln\left(\frac{1-\rho_0\delta_2/\delta_1}{1-\rho_0\delta_1/\delta_2}\right)\right\}\right]$$

where χ_2^{2n} is a chi-squared random variable with 2n degrees of freedom. The power function is monotone decreasing in $C(\alpha)$ so that critical values can be found for any level of significance.
• Bhattacharyya and Johnson (1973) show that the level uniformly most powerful test (UMP) of $H_0 : \lambda_{12} = 0$ versus $H_1 : \lambda_{12} > 0$ in the symmetric MOBVE model with $\lambda_1 = \lambda_2$ is given by: reject H_0 if $\lambda_{12} > 0$ or if $2S_1/(S_2 - S_1) > F_{2n,2n,\alpha}$, the upper α percentage point of a central F distribution with (2n,2n) degrees of freedom.

3.4 Bayesian inference for the Marshall-Olkin Model

Pena and Gupta (1990) have provided a comprehensive account of Bayesian estimation of Marshall Olkin's distribution under series and parallel sampling.
Let N_1, N_2, N_{12}, T defined in the inference of MOBVE and
$W = \sum_{i=1}^{n} \delta_j (T'')_i, j = 1, 2$ with $T'' = \max(X, Y) - \min(X, Y)$, $\delta_1 = I(X > Y)$ and $\delta_2 = I(X < Y)$

3.4.1 Series sampling

The joint density

Let $\Theta = (\lambda_1, \lambda_2, \lambda_{12})$
Using the fact that the marginals and the minimum are exponential, and the distribution of I(X>Y), I(X>Y)are multinomial, the joint distribution in series sampling is given by

$$f_s(\min(X, Y), \delta_1, \delta_2|\Theta) = \lambda \exp(-\lambda \min(X, Y))(\lambda_1/\lambda)^{n_1}(\lambda_2/\lambda)^{n_2}(1 - \lambda_1/\lambda - \lambda_2/\lambda)^{n_{12}}, \tag{3.2}$$

with n_1, n_2, n_{12} the realisation of (X>Y) (X<Y) and (X=Y) respectively
From the joint density it is clear that (N_1, N_2, T) are a sufficient statistics for Θ under series sampling.

Prior distribution

They start with $\lambda = \lambda_1 + \lambda_2 + \lambda_{12}$, it is the rate of the Poisson process that generates all the shocks. Plausible family of priors for λ the gamma family with generic element

$$\pi(\lambda) \propto \lambda^{\alpha-1} \exp(-\beta\lambda), \qquad \alpha > 0, \beta > 0$$

For the ratio λ_i/λ given λ they proposed as family of priors the Dirichlet family with generic element (The Drichelet distribution is often used for the parameters of multinomial distribution)

$$\pi(\lambda_1/\lambda, \lambda_2/\lambda|\lambda) \propto (\lambda_1/\lambda)^{\alpha_1(\lambda)-1}(\lambda_2/\lambda)^{\alpha_2(\lambda)-1}(1 - \lambda_1/\lambda - \lambda_2/\lambda)^{\alpha_{12}(\lambda)-1},$$

where $\alpha_i(\lambda) > 0$, authors have assumed in their paper that the α_{is} do not depend on λ. Thus, the density for $(\lambda, \lambda_1/\lambda, \lambda_2/\lambda)$ belongs to the Gamma-Dirichlet family and is of the form

$$\pi_s(\lambda, \tfrac{\lambda_1}{\lambda}, \tfrac{\lambda_2}{\lambda}|\beta, \alpha, \alpha_1, \alpha_2, \alpha_{12})$$
$$= \tfrac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}\exp(-\beta\lambda)\tfrac{1}{B(\alpha_1,\alpha_2,\alpha_{12})}\left(\tfrac{\lambda_1}{\lambda}\right)^{\alpha_1-1}\left(\tfrac{\lambda_2}{\lambda}\right)^{\alpha_2-1}\left(1-\tfrac{\lambda_1}{\lambda}-\tfrac{\lambda_2}{\lambda}\right)^{\alpha_{12}-1} \quad (3.3)$$

for $\lambda \geq 0, \lambda_i/\lambda \geq 0, 0 \leq \lambda_1/\lambda + \lambda_2/\lambda \leq 1$, and where

$$B(\alpha_1, \alpha_2, \alpha_{12}) = \prod_{i=0}^{2} \Gamma(\alpha_i)/\Gamma(\Delta),$$

$\Delta = \alpha_1 + \alpha_1 + \alpha_{12}$, and $\Gamma(.)$ is gamma function.
The function in equation (3.3)is denoted by $GD(\beta, \alpha, \alpha_1, \alpha_2, \alpha_{12})$.
It is interest to note that Jeffreys's non-informative prior models prior dependence among the $\lambda_i s$ which is a limiting case of equation (3.3) and it is given by

$$GD(0, 0, \frac{1}{2}, \frac{1}{2}, \frac{1}{2})$$

Theorem
if Θ has prior distribution $GD(\beta, \alpha, \alpha_1, \alpha_2, \alpha_{12})$.then
$\star$ $E(\lambda) = \alpha/\beta$ and $V(\lambda) = \alpha/\beta^2$
$\star$ $E(\lambda_i) = (\alpha/\Delta)(\alpha_i/\beta), \quad i = 1, 2, 12$
$\star$ $V(\lambda_i) = (\alpha/\Delta)(\alpha_i/\beta^2)\{(\alpha_i + 1)(\alpha + 1)/(\Delta + 1) - \alpha_i\alpha/\Delta\}$,
$\star$ $Cov(\lambda_i, \lambda_j) = (\alpha_i\alpha_j/\beta^2)(\alpha/\Delta))(\alpha + 1)/(\Delta + 1), i \neq j$
$\star$the prior modes of λ and λ_i are given respectively by $(\alpha - 3)/\Delta$ and $\{(\alpha - 3)/(\Delta - 3)\}\{(\alpha_i - 1/\beta\}$, provided that $\alpha > 3$ and $\alpha_i > 1$.

Bayes Estimator of Θ

Suppose we have a sample in series of size n and let (t, n_1, n_2) be the corresponding realization of the minimal sufficient statistic (T, N_1, N_2). From equations (3.2) and (3.3) and by Bayes theorem, the posterior distribution of $(\lambda, \lambda/\lambda_1, \lambda_2/\lambda)$, given $(t, n_1 > n_2)$, is

$$GD(\beta + t, \alpha + n, \alpha_{12} + n_{12}, \alpha_1 + n_1, \alpha_2 + n_2), \quad (3.4)$$

Under the quadratic loss function the estimator of $\widehat{\Theta}_s = (\widehat{\lambda}_{1s}, \widehat{\lambda}_{2s}, \widehat{\lambda}_{12s})$, is the mean of the posterior distribution (3.4).using the previous theorem this estimators are given by

$$\begin{aligned} \widehat{\lambda}_{is} &= \frac{\alpha+n}{\Delta+n}\frac{\alpha_i+N_i}{\beta+T} \\ &= \frac{\Delta}{\Delta+n}\frac{\alpha+n}{\alpha}\frac{\beta}{\beta+T}\frac{\alpha}{\Delta}\frac{\alpha_i}{\beta} + \frac{\alpha+n}{\Delta+n}\frac{T}{\beta+T}\frac{N_i}{T}, \end{aligned} \tag{3.5}$$

which is a linear combination of the prior mean of λ_i and the MLE of λ_i given by $\check{\lambda}_i = N_i/T$ (see Proschan and Sullo (1974)).
And $\lim(\widehat{\lambda}_1, \widehat{\lambda}_2, \widehat{\lambda}_{12}) = (\check{\lambda}_1, \check{\lambda}_2, \check{\lambda}_{12})$ when β, α, α_1, α_2, and α_{12} go to zero.
A referee suggested that another estimate of Θ is the posterior mode which is given by

$$\widetilde{\lambda}_{is} = \frac{\alpha+n-3}{\Delta+n-3}\frac{\alpha_i+N_i-1}{\beta+T}$$

provided that $\alpha + n > 3$ and $\alpha_i + n_i > 1$ with $i = 1, 2, 12$

3.4.2 Parallel sampling

The joint density

the joint density in parallel sampling is given by
$f_p(\min(x, y), \delta_1, \delta_2, t''|\Theta)$

$$= f_s(\min(x, y), \delta_1, \delta_2|\Theta)\Pi_{i=1}^2[(\lambda - \lambda_i)^{n_i}\exp(-(\lambda - \lambda_i)w\delta_i)]\{I(w = 0)\}^{n12} \tag{3.6}$$

(T, N_1, N_2, W_1, W_2) are the minimal sufficient statistics for Θ under parallel sampling.

Prior and Posterior distribution for Θ

The authors have suggested the same family of prior distribution chosen in series sampling given in equation (3.3). Justifying this choice in the fact that no additional prior information about Θ can be gained by knowing that the components are connected in parallel rather than series.

From the formula of the joint density in parallel sampling (3.6) and this family of prior distribution given in (3.4), we can seen that, this family priors does not correspond to the conjugate family in this case.
They do not use the advantage of limit case of prior family which corresponds to Jeffreys's no informative prior, because they are not certain whether the choice of $\beta = \alpha = \alpha_1 = \alpha_2 = \alpha_{12} = 0$ leads to the MLEs under parallel sampling.
Let (t, n_1, n_2, w_1, w_2) be the realization of the minimal sufficient (T, N_1, N_2, W_2, W_2).the posterior distribution is given by

$$\begin{aligned}\pi_p(\lambda, \frac{\lambda_1}{\lambda}, \frac{\lambda_2}{\lambda}|t, n_1, n_2, w_1, w_2) \quad &\propto \quad \lambda^{\alpha^*} \exp(-\beta^* \lambda)(\lambda_1/\lambda)^{\alpha_1^*-1}(\lambda_2/\lambda)^{\alpha_2^*-1}(1-\lambda_1/\lambda-\lambda_2/\lambda)^{\alpha_{12}^*-1} \\ &\times \quad (1-\lambda_1/\lambda)^{n_1}(1-\lambda_2/\lambda)^{n_2} \exp[-\lambda\{(1-\lambda_1/\lambda)s_1+(1-\lambda_2/\lambda)s_2\}] \quad (3.7)\end{aligned}$$

With respect to the quadratic loss function, the estimators of Bayes are

$$\widehat{\lambda}_{1p} = \frac{A(\beta+T, \alpha+n+N_1+N_2, \alpha_{12}+N_{12}, \alpha_1+N_1, \alpha_2+N_2, N_1, N_2, W_1, W_2)}{A(\beta+T, \alpha+n+N_1+N_2+1, \alpha_{12}+N_{12}+1, \alpha_1+N_1, \alpha_2+N_2, N_1, N_2, W_1, W_2)}$$

$$\widehat{\lambda}_{2p} = \frac{A(\beta+T, \alpha+n+N_1+N_2, \alpha_{12}+N_{12}, \alpha_1+N_1, \alpha_2+N_2, N_1, N_2, W_1, W_2)}{A(\beta+T, \alpha+n+N_1+N_2+1, \alpha_{12}+N_{12}, \alpha_1+N_1+1, \alpha_2+N_2, N_1, N_2, W_1, W_2)}$$

and

$$\widehat{\lambda}_{3p} = \frac{A(\beta+T, \alpha+n+N_1+N_2, \alpha_{12}+N_{12}, \alpha_1+N_1, \alpha_2+N_2, N_1, N_2, W_1, W_2)}{A(\beta+T, \alpha+n+N_1+N_2+1, \alpha_{12}+N_{12}, \alpha_1+N_1, \alpha_2+N_2+1, N_1, N_2, W_1, W_2)}$$

with $A(\beta^*, \alpha^*, \alpha_1^*, \alpha_2^*, \alpha_{12}^*, n_1, n_2, s_1, s_1)$: is the proportionality constant needed to make expression (3.7) a density function.
The estimators in parallel sampling are not in closed form.
• A measure of the precision of the Bayes estimates is their posteriors variances under series sampling and its matrix variance-covariance matrix under parallel sampling.

3.4.3 Gain in precision of parallel over series sampling

In practice, estimation of parameters using a series sample is easier compared with that for a parallel sample because it is more difficult to compute parallel sample estimates, and it takes more experimental time to obtain the parallel sample.

Pena and Gupta have examined the gain in precision of parallel sample estimates over series sample estimates the asymptotic normality of the highest posterior density under series and parallel sampling.

The meausue of asymptotic efficiency of $\widehat{\lambda}_{ip}$ over $\widehat{\lambda}_{is}$ is the ratio of their asymptotic variances given by

$$eff(\widehat{\lambda}_{is} : \widehat{\lambda}_{ip}) = avar(\widehat{\lambda}_{is} n^{1/2})/avar(\widehat{\lambda}_{ip} n^{1/2}), \quad i = (1, 2, 12)$$

These asymptotic variances are the diagonal elements of the information matrix of the joints densities under parallel ans series sampling , and the fact that the MLEs and Bayes estimators are consistent.

Theorem

The asymptotic efficiencies of parallel sample estimates over series sample estimates are given by

$$eff(\widehat{\lambda}_{1s} : \widehat{\lambda}_{1p}) = 1 + \frac{\lambda\lambda_1\lambda_2(\lambda_2 + \lambda_{12})}{\lambda(\lambda_1 + \lambda_{12})^2(\lambda_2 + \lambda_{12}) + \lambda_2\lambda_{12}(\lambda_2 + \lambda_{12})^2\lambda_1\lambda_2^2\lambda_{12}}$$

$$eff(\widehat{\lambda}_{2s} : \widehat{\lambda}_{2p}) = 1 + \frac{\lambda\lambda_1\lambda_2(\lambda_2 + \lambda_{12})}{\lambda(\lambda_1 + \lambda_{12})^2(\lambda_2 + \lambda_{12})^2 + \lambda_2\lambda_{12}(\lambda_1 + \lambda_{12})^2\lambda_1^2\lambda_2\lambda_{12}}$$

$$eff(\widehat{\lambda}_{12s} : \widehat{\lambda}_{12p}) = 1 + \frac{\lambda_1\lambda_{12}}{(\lambda_2\lambda_2)^{12} + \lambda_1\lambda_2} \frac{\lambda_2\lambda_{12}}{(\lambda_1\lambda_2)^{12} + \lambda_1\lambda_2}$$

As a special case this theorem , if $\lambda_1 = \lambda_2 = \lambda_{12}$ then $eff(\widehat{\lambda}_{12s} : \widehat{\lambda}_{12p}) = 7/5$ and $eff(\widehat{\lambda}_{is} : \widehat{\lambda}_{ip}) = 35/29 (i = 1, 2)$. Parallel sample estimates could therefore be considerably more efficient than series sample estimates.

3.5 Some recent works on Bayesian analysis of bivariate distributions

3.5.1 Objective Bayesian analysis for bivariate Marshall-Olkin exponential distribution

Pena and Gupta (1990) have proposed the informative (conjugate) Bayesian analysis for bivariate Marshall-Olkin exponential distribution but the obtained estimators under parallel sampling are not in closed forms. Qiang Guan, Yincai Tang and Ancha Xu (2013) have proposed the non informative (objective) Bayesian analysis for this model. They have assumed two type of the non informative priors: reference priors and matching priors. They have used Gibbs sampling to obtain the credible intervals and coverage probabilities of parameters. Closed forms of Bayesian estimators are obtained with respect to the quadratic loss function. They applied their procedures to a real data set. The result show that the Bayesian estimates perform better than MLE.

3.5.2 Marshall-Olkin bivariate Weibull distribution

Marshall and Olkin (1967) have extended the MOBE model to the Marshall-Olkin bivariate Weibull (MOBW) model.
If $(X_1, X_2) \sim \text{MOBW}(\alpha, \lambda_0, \lambda_1, \lambda_2)$, the joint SF of (X_1, X_2) can be written as

$$\begin{aligned} S_{X_1,X_2}(x_1,x_2) &= P(X_1 > x_1, X_2 > x_2) = P(U_1 > x_1, U_2 > x_2, U_0 > \max\{x_1,x_2\}) \\ &= S_{WE}(x_1;\alpha,\lambda_1)S_{WE}(x_2;\alpha,\lambda_2)S_{WE}(\max\{x_1,x_2\};\alpha,\lambda_0) \end{aligned}$$

$$S_{X_1,X_2}(x_1,x_2) = \begin{cases} S_{WE}(x_1;\alpha,\lambda_1)S_{WE}(x_2;\alpha,\lambda_0+\lambda_2), & \text{if } x_1 < x_2 \\ S_{WE}(x_1;\alpha,\lambda_0+\lambda_1)S_{WE}(x_2;\alpha,\lambda_2), & \text{if } x_1 > x_2 \\ S_{WE}(x;\alpha,\lambda_0+\lambda_1+\lambda_2), & \text{if } x = x_1 = x_2 \end{cases}$$

Therefore, the joint PDF of X_1 and X_2 can be obtained as

$$f_{X_1,X_2}(x_1,x_2) = \begin{cases} f_1(x_1,x_2), & \text{if } x_1 < x_2 \\ f_2(x_1,x_2), & \text{if } x_1 > x_2 \\ f_0(x), & \text{if } x = x_1 = x_2 \end{cases}$$

where

$$f_1(x_1,x_2) = f_{WE}(x_1;\alpha,\lambda_1)f_{WE}(x_2;\alpha,\lambda_0+\lambda_2),$$

$$f_2(x_1,x_2) = f_{WE}(x_1;\alpha,\lambda_0+\lambda_1)f_{WE}(x_2;\alpha,\lambda_2,$$

$$f_0(x) = \frac{\lambda_0}{\lambda_0+\lambda_1+\lambda_2}f_{WE}(x;\alpha,\lambda_0+\lambda_1+\lambda_2),$$

•Kundu and Dey (2009) proposed the EM algorithm to estimate the parameters of the MOBW distribution, where they used the bivariate observed data rather than competing risks data, but it is well known that the MLEs do not always exist.

• Debasis Kundu et al (2013) have proposed the Bayesian analysis of the unknown parameters of Marshall-Olkin Bivariate Weibull Distribution. This analysis is closely related to the analysis given by Pena and Gupta (1990) for estimate the parameters of MOBVE under series and parallel sampling.

This study is done in two phases when the shape parameter is known and when it is unknown. They have not used any specific form of prior on the shape parameter, they have only assumed that the PDF of the prior distribution is log-concave on $[0, \infty[$.

For the scale parameters, they have used the same Gamma-Dirichel prior proposed by Pena and Gupta (1990). they have used the squared error as loss function. The Bayes estimators cannot be obtained in explicit forms. The authors have used the importance sampling technique to compute the Bayes estimators and also compute the associated HPD credible intervals.

They have performed the analysis of a data set mainly for illustrative purposes and it is observed that the proposed method work quite well, in fact it is observed that the Bayes estimators exist even when MLEs do not exist. More, with informative priors, the Bayes estimators perform better than the MLEs.

3.5.3 Bayesian analysis of three parameter absolute continuous Marshall-Olkin bivariate Pareto distribution

Assume that U_0, U_1, and U_2 are three independent random variables where:

$$U_0 \sim PA(II)(0,1,\alpha_0)\ , U_1 \sim PA(II)(\mu_1,\sigma_1,\alpha_1)\ , U_2 \sim PA(II)(\mu_2,\sigma_2,\alpha_2)$$

We define $X_1 = \min\{\sigma_1 U_0 + \mu_1, U_1\}$ and $X_2 = \min\{\sigma_2 U_0 + \mu_2, U_2\}$, then (X_1, X_2) has the Marshall-Olkin bivariate Pareto (MOBVPA) distribution of second kind, we call it as $BVPA(\mu_1,\mu_2,\sigma_1,\sigma_2,\alpha_0,\alpha_1,\alpha_2)$.
the survival function of (X_1, X_2) can be written for $z = \max\{\frac{x_1-\mu_1}{\sigma_1}, \frac{x_2-\mu_2}{\sigma_2}\}$

$$S_{X_1,X_2}(x_1,x_2) = (1+z)^{-\alpha_0}(1+\frac{x_1-\mu_1}{\sigma_1})^{-\alpha_1}(1+\frac{x_2-\mu_2}{\sigma_2})^{-\alpha_2}$$

$$S_{X_1,X_2}(x_1,x_2) = \begin{cases} S_1(x_1,x_2), & if\ \frac{x_1-\mu_1}{\sigma_1} < \frac{x_2-\mu_2}{\sigma_2} \\ S_2(x_1x_1,x_2), & if\ \frac{x_1-\mu_1}{\sigma_1} > \frac{x_2-\mu_2}{\sigma_2} \\ S_0(x), & if\ \frac{x_1-\mu_1}{\sigma_1} = \frac{x_2-\mu_2}{\sigma_2}\ and\ x = x_1 = x_2 \end{cases}$$

where

$$S_1(x_1,x_2) = (1+\frac{x_2-\mu_2}{\sigma_2})^{-\alpha_0-\alpha_2}(1+\frac{x_1-\mu_1}{\sigma_1})^{-\alpha_1},$$

$$S_2(x_1,x_2) = (1+\frac{x_1-\mu_1}{\sigma_1})^{-\alpha_0-\alpha_1}(1+\frac{x_2-\mu_2}{\sigma_2})^{-\alpha_2},$$

$$S_0(x) = (1+\frac{x-\mu_1}{\sigma_1})^{-\alpha_0-\alpha_1-\alpha_2},$$

The joint distribution is given by

$$f_{X_1,X_2}(x_1,x_2) = \begin{cases} f_1(x_1,x_2), & \text{if } \frac{x_1-\mu_1}{\sigma_1} < \frac{x_2-\mu_2}{\sigma_2} \\ f_2(x_1x_1,x_2), & \text{if } \frac{x_1-\mu_1}{\sigma_1} > \frac{x_2-\mu_2}{\sigma_2} \\ f_0(x), & \text{if } \frac{x_1-\mu_1}{\sigma_1} = \frac{x_2-\mu_2}{\sigma_2}\ and\ x = x_1 = x_2 \end{cases}$$

where

$$f_1(x_1,x_2) = \frac{\alpha_1}{\sigma_1\sigma_2}(\alpha_0+\alpha_2)(1+\frac{x_2-\mu_2}{\sigma_2})^{-(\alpha_0+\alpha_2+1)}(1+\frac{x_1-\mu_1}{\sigma_1})^{-(\alpha_1+1)},$$

$$f_2(x_1, x_2) = \frac{\alpha_2}{\sigma_1\sigma_2}(\alpha_0 + \alpha_1)(1 + \frac{x_1 - \mu_1}{\sigma_1})^{-(\alpha_0+\alpha_1+1)}(1 + \frac{x_2 - \mu_2}{\sigma_2})^{-(\alpha_2+1)},$$
$$f_0(x) = \alpha_0(1 + x)^{-(\alpha_0+\alpha_1+\alpha_2+1)},$$

• To estimate the parameters of the MOBVPA distribution the EM algorithm are proposed by several authors. Recently, Asimit, Furman and Vernic (2016) used the EM algorithm to estimate the singular Marshall-Olkin bivariate Pareto distribution. However, there is no paper available for formulation and estimation of the absolute continuous version of this distribution. A few recent works as Kundu and Gupta (2009), Kundu and Gupta (2010) and Kundu, Kumar and Gupta (2015) include the Marshall-Olkin bivariate distribtion. Estimation through the EM algorithm can been seen there. But in the estimation of the absolute continuous version of bivariate Pareto distribution through the EM algorithm is not straight forward. The problem becomes more complicated when we consider the location and scale parameter in our problem.

• Biplap Paul, Debasis Kundu et al (2018) consider the Bayesian analysis of absolute continuous version of Marshall-Olkin bivariate Pareto distribution whose marginals are not Type II univariate Pareto distributions. They use the notation BB-BVPA for absolute continuous Marshall-Olkin bivariate Pareto. This form of Marshall-Olkin bivariate Pareto is similar to absolute continuous bivariate exponential distribution as proposed by Block and Basu (1974).

For the prior distribution, they assume that α_0, α_1, and α_2 are distributed according to the gamma distribution with shape parameter k_i and scale parameter θ_i. They have taken reference prior as prior for each parameter conditional on the others.The Bayes estimator cannot be obtained in closed form. They propose to use two approach Lindley approximation and Slice cum Gibbs sampler. However, we can use other Monte Carlo methods for the same. They use EM algorithms instead of MLE. Bayesian estimates of the parameters of absolute continuous bivariate Pareto under square error loss. Credible intervals are also provided for all methods and all prior distributions. A real-life data analysis is shown for illustrative purpose.

3.5.4 Bayesian analysis of absolute continuous Marshall-Olkin bivariate Pareto distribution with location and scale parameters

Biplab Paul et al(2018) have proposed Bayesian analysis of absolute continuous Marshall-Olkin bivariate Pareto distribution with location and scale parameters. They have taken gamma prior for shape and scale parameters and truncated normal for location parameters. They propose two different novel approaches of slice sampling to estimate the parameters of absolute continuous Marshall-Olkin bivariate Pareto distribution with location and scale parameters.

Approach 1: In this approach they propose to use usual step out methods for slice sampling for each conditional distributions. Standard slice sampling works only for continuous posterior.

Approach 2: In this approach they use directly discontinuous conditional posterior based on likelihood of bivariate distribution.

• Credible intervals and coverage probabilities are also provided for all methods. A real-life data analysis is shown for illustrative purpose. Two approaches proposed works quite well even for moderately large sample size. However coverage probabilities calculated based on parametric bootstrap samples show that the proposed credible interval construction does not work very well for all the shape parameters. Although we see its significant improvement for large sample, more research is needed to explore the best credible interval construction in small sample set up. The same study can be made using many other algorithms like importance sampling,.etc.

3.5.5 Comparison of the maximum likelihood and Bayes estimators for symmetric bivariate exponential distribution under different loss function (Chadli et al, 2013)

In this paper, authors are interested to the estimation of the parameters and time between failure of the bivariate exponential distribution based on the bivariate exponential distribution of Freund (1961) and Block and Basu (1974) given in section (2.3.1). Using Monte Carlo study, they perform the maximum likelihood approach and the Bayesian methodology and compare the estimators with respect th various loss function

The model used in this work is

$$f(x,y) = \begin{cases} \theta\lambda'^2 \exp\{-2\lambda' x - \theta\lambda'(y-x)\} & \text{if } \ 0 < x < y \\ \theta\lambda'^2 \exp\{-2\lambda' y - \theta\lambda'(x-y)\} & \text{if } \ 0 < x < y \end{cases}$$

By assuming in the Freund model identical components, $\alpha = \beta = \lambda'$ and the reparametrization of the Block Basu bivariate exponential distribution by assuming $\lambda_1 = \lambda_2 = \lambda'(2-\theta)$ and $\lambda_{12} = 2\lambda'(\theta - 1)$

• In first time , authors have considered the maximum likelihood, then the Bayesian approach to estimate the parameters and time between failure using conjugate and no informative prior. They have considered different loss functions (squared loss, absolute loss, deGroot loss, linex loss and Entropy loss).

• They have conclude that the maximum likelihood estimation can be improved by Bayesian method, and they prove that this improvement can be efficient (Pitman closeness criterion and the relative efficiency) using suitable loss function.

3.6 Simulation study

In this section, we verify the performance of the proposed Bayesian method, and we compare them with the MLEs obtained by EM algorithm given by Kundu and Dey(2009). As it is observed that the Bayesian estimators cannot be obtained easily in explicit forms in general, we propose to use the importance sampling to compute the Bayesian estimators. In first time, we simulate the bivariate Marshall-Olkin exponential distribution. We use the programming software R to perform our simulations.

3.6.1 Simulation of the bivariate Marshall-Olkin exponential distribution

The sample data are generated based on following algorithm :

• Step 1: Generate z_i using the exponential distribution with failure rate λ_i for i=1,2,12.

• Step 2: Take $x = \min(z_1, z_{12})$ and $y = \min(z_2, z_{12})$ and, therefore, (X, Y) follows a bivariate exponential distribution of Marshall-Olkin type.

x		y	
Min.	:0.000086	Min.	:0.000086
1st Qu.	:0.132447	1st Qu.	:0.109000
Median	:0.325063	Median	:0.260927
Mean	:0.480426	Mean	:0.377591
3rd Qu.	:0.669446	3rd Qu.	:0.520669
Max.	:4.599179	Max.	:3.198889

Table 3.2: The summary of realization

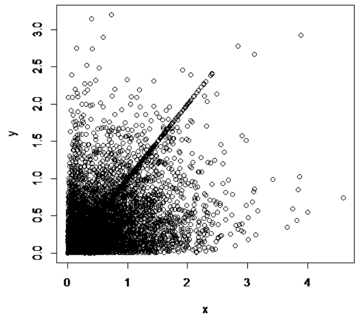

Figure 3.2: The graph of realization

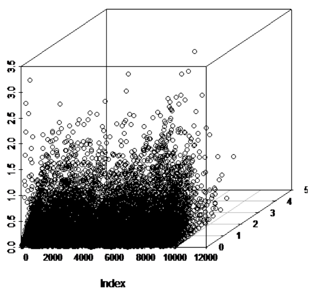

Figure 3.3: Three dimensional density plots of realization MOBVE

• From figure 3.2 and 3.3, we can see that the Marshall-Olkin bivariate exponential distribution is not continuous (lines x=y).

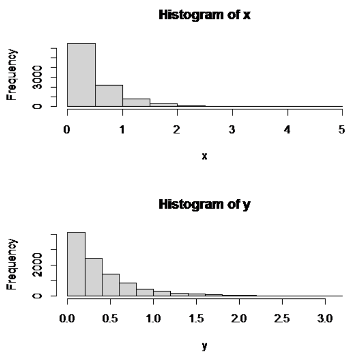

Figure 3.4: Histogram of marginal

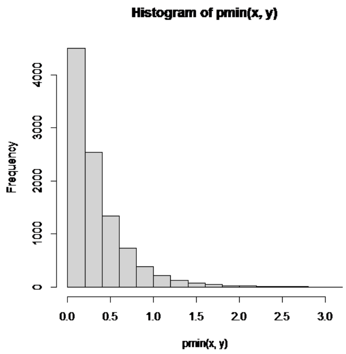

Figure 3.5: Histogram of the minimum

• From figure 3.4 and 3.5, it is clear that the marginal distributions and the distribution of the minimum are exponential.

• To confirm that the simulated data really follow MOBVE, we perform a Kolmogorov-Smirnov fit test for the min(x, y) (if it follow the exponential distribution)

```
> ks.test(pmin(x,y),pexp,4)

        One-sample Kolmogorov-Smirnov test

data:  pmin(x, y)
D = 0.0061778, p-value = 0.8399
alternative hypothesis: two-sided
```

Figure 3.6: The Kolmogorov-Smirnov fit test

• From the p-value=0.8399, we cannot reject the hypotheses that min(x, y) follow exponential distribution, so the simulated data follow MOBVE.

3.6.2 Importance sampling

In the previous sections, it is observed that the Bayesian estimators cannot be obtained easily in explicit forms, Kundu and Gupta (2013) have proposed the importance sampling procedure to obtain the Bayesian estimators of the parameters of bivariate Marshall-Olkin weibull distribution. We will use the same procedure to obtain the Bayesian estimators of the parameters of bivariate Marshall-Olkin exponential distribution.
• Bayesian estimators obtained are based on likelihood function given below :

$$L(x, y|\lambda_1, \lambda_2, \lambda_{12}) = \Pi_{x<y} f_1(x, y) \Pi_{x>y} f_2(x, y) \Pi_{x=y} f_3(x, y)$$

with

$$f_1(x,y) = \begin{cases} \lambda(\lambda_2+\lambda_{12})e^{-\lambda_1 x-(\lambda_2+\lambda_{12})y}, & 0<x<y<\infty \\ 0, & \text{otherwise} \end{cases}$$

$$f_2(x,y) = \begin{cases} \lambda(\lambda_1+\lambda_{12})e^{-(\lambda_1+\lambda_{12})x-\lambda_2 y}, & 0<y<x<\infty \\ 0, & \text{otherwise} \end{cases}$$

and

$$f_3(x,y) = \begin{cases} \lambda e^{-\lambda x}, & 0<y=x<\infty \\ 0, & \text{otherwise} \end{cases}$$

• Unlike Kundu and Gupta (2013), we assume that the parameters are independent, and that the prior of all the parameters are Gamma distributions, that is:

$$\pi_1(\lambda_1|a_1,b_1) = \frac{b_1^{a_1}}{\Gamma(a_1)}\lambda_1^{a_1}e^{-b_1\lambda_1}$$

$$\pi_2(\lambda_2|a_2,b_2) = \frac{b_2^{a_2}}{\Gamma(a_2)}\lambda_2^{a_2}e^{-b_2\lambda_2}$$

$$\pi_3(\lambda_{12}|a_3,b_3) = \frac{b_3^{a_3}}{\Gamma(a_3)}\lambda_{12}^{a_3}e^{-b_3\lambda_{12}}$$

• Based on the likelihood function and the priors distribution, the joint posterior density function of $(\lambda_1, \lambda_2, \lambda_{12})$ can we written as

$$\begin{aligned} \pi(\lambda_1,\lambda_2,\lambda_{12}|x,y) &= \pi_1(\lambda_1|a_1,b_1)\pi_2(\lambda_2|a_2,b_2)\pi_3(\lambda_{12}|a_3,b_3)L(x,y|\lambda_1,\lambda_2,\lambda_{12}) \\ &\propto (\lambda_2+\lambda_{12})^{n_1}(\lambda_1+\lambda_{12})^{n_2} \times Gamma(\lambda_{12};a_3+n_3,T_3+b_3) \\ &\times\ Gamma(\lambda_1;a_1+n_1,T_1+b_1) \times Gamma(\lambda_2;a_2+n_2,T_2+b_2) \end{aligned}$$

where

$$T_1 = \sum_{x>y \bigcup x<y} x_i + \sum_{x=y}(x_i,y_i)$$

$$T_2 = \sum_{y>x \bigcup y<x} y_i + \sum_{x=y}(x_i,y_i)$$

$$T_3 = \sum_{x>y} x + \sum_{y>x} y_i + \sum_{x=y} (x_i, y_i)$$

• Denote that $\theta = (\lambda_1, \lambda_2, \lambda_{12})$. Then the Bayesian estimator of θ under the squared error loss function is

$$\hat{\theta}_B = \frac{\int_0^\infty \int_0^\infty \int_0^\infty \theta L(\lambda_1, \lambda_2, \lambda_{12}|x, y) d\lambda_1 d\lambda_2 d\lambda_{12}}{\int_0^\infty \int_0^\infty \int_0^\infty L(\lambda_1, \lambda_2, \lambda_{12}|x, y) d\lambda_1 d\lambda_2 d\lambda_{12}}$$

• We use the following importance sampling procedure to obtain $\hat{\theta}_B$:

• Step 1: Generate

$$\lambda_1 \sim Gamma(\lambda_1; a_1 + n_1, T_1 + b_1)$$

$$\lambda_2 \sim Gamma(\lambda_2; a_2 + n_2, T_2 + b_2)$$

$$\lambda_{12} \sim Gamma(\lambda_3; a_3 + n_3, T_3 + b_3)$$

• Step 2: Repeat this procedure to obtain $\{(\lambda_{1i}, \lambda_{2i}, \lambda_{12i}); i = 1....n\}$.

• Step 3: the approximate value of the estimator can be obtained as:

$$\frac{\sum_{i=1}^n \theta_i h(\lambda_{1i}, \lambda_{2i}, \lambda_{12i})}{\sum_{i=1}^n h(\lambda_{1i}, \lambda_{2i}, \lambda_{12i}}$$

with $h(\lambda_1, \lambda_2, \lambda_{12}) = (\lambda_1 + \lambda_{12})^{n_2} (\lambda_2 + \lambda_{12})^{n_1}$

3.6.3 Numerical simulation

In the process of simulation, we assume that the prior distribution of λ_1, λ_2 and λ_{12} are $Gamma(1, 1)$ and we take $(\lambda_1, \lambda_2, \lambda_{12}) = (1, 1, 1)$ in order to compare the results with the results obtained by Kundu and Dey (2009). We take our sample size as n=25, 50,100.

• The results are listed in table below:

n	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\lambda}_{12}$
25	0.677	0.677	0.705
50	0.771	0.771	0.79
100	0.843	0.843	0.855

Table 3.3: The Bayesian estimators of $(\lambda_1, \lambda_2, \lambda_{12}) = (1, 1, 1)$

• The results obtained by EM algorithm given by Kundu and Dey (2009) are listed in table below

n	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\lambda}_{12}$
25	0.646	0.675	0.666
50	0.761	0.747	0.771
100	0.841	0.839	0.843

Table 3.4: The maximum likelihood estimators of $(\lambda_1, \lambda_2, \lambda_{12}) = (1, 1, 1)$

Important remarks

1. Both Bayesian method and maximum likelihood method perform better when the sample size increases.
2. When the sample size is small or moderate, the parameters based on the Bayesian method are better than these based on EM algorithm.

3.7 Real data application

The data set presented in Table 3.4 is from Meintanis (2007). The data represent the football (soccer) data where at least one goal scored by the home team and at least one goal scored directly from a penalty kick, foul kick or any other direct kick (all of them together will be called as kick goal) by any team. Here X_1 represents the time in minutes of the first kick goal scored by any team and X_2 represents the first goal of any type scored by the home team. In this case all possibilities are open, for example $X_1 < X_2$, or $X_1 > X_2$, or $X_1 = X_2$.

2005-2006	X_1	X_2	2004-2005	X_1	X_2
Lyon-Real Madrid	26	30	nternazionale-Bremen	34	34
Milan-Fenerbahce	63	18	Real Madrid-Roma	53	39
Chelsea-Anderlecht	19	19	Man. United-Fenerbahce	54	7
Club Brugge-Juventus	66	85	Bayern-Ajax	51	28
Fenerbahce-PSV	40	40	Moscow-PSG	76	64
Internazionale-Rangers	49	49	Barcelona-Shakhta	64	15
Panathinaikos-Bremen	8	8	Leverkusen-Roma	26	48
Ajax-Arsenal	69	71	Arsenal-Panathinaikos	16	16
Man. United-Benfica	39	39	Dynamo Kyiv-Real Madrid	44	13
Real Madrid-Rosenborg	82	48	Man. United-Sparta	25	14
Villarreal-Benfica	72	72	Bayern-M. TelAviv	55	11
Juventus-Bayern	66	62	Bremen-Internazionale	49	49
Club Brugge-Rapid	25	9	Anderlecht-Valencia	24	24
Olympiacos-Lyon	41	3	Panathinaikos-PSV	44	30
Internazionale-Porto	16	75	Arsenal-Rosenborg	42	3
Schalke-PSV	18	18	Liverpool-Olympiacos	27	47
Barcelona-Bremen	22	14	M. Tel-Aviv-Juventus	28	28
Milan-Schalke	42	42	Bremen-Panathinaikos	2	2
Rapid-Juventus	36	52			

Table 3.5: UEFA Champions League data

• Kundu and Gupta (2013), Kundu and Dey (2009) have suggested MOBW to fit the data, which is given by:

$$S_{X_1,X_2}(x_1,x_2) = \begin{cases} S_{WE}(x_1;\alpha,\lambda_1)S_{WE}(x_2;\alpha,\lambda_0+\lambda_2), & \text{if } x_1 < x_2 \\ S_{WE}(x_1;\alpha,\lambda_0+\lambda_1)S_{WE}(x_2;\alpha,\lambda_2), & \text{if } x_1 > x_2 \\ S_{WE}(x;\alpha,\lambda_0+\lambda_1+\lambda_2), & \text{if } x = x_1 = x_2 \end{cases}$$

• Kundu and Gupta have proposed Bayesian method to find the point estimates. In this case n=37,$n_1 = 6$, $n_2 = 17$, $n_3 = 14$. They have assumed that the prior distributions of λ_0, λ_1 and λ_2 are $Gamma(1,1)$. And the prior of α is $Gamma(0.001, 0.001)$. Using importance sampling, and with respecting to squared loss function they have computed a consistent Bayes estimators of α ,λ_0, λ_1 and λ_2 and their corresponding credible intervals:

unknown parameters	Bayes estimator	credible interval
α	1.7049	[1.3280,1.8920]
λ_0	2.075	[1.4852,2.3571]
λ_1	0.9566	[0.5045,1.4816]
λ_2	3.368	[2.1901,3.5573]

Table 3.6: Bayes estimator and credible interval of the unknown parameters

• Now a natural question is how good is fit. Unfortunately, we do not any proper bivariate goodness of the fit test for general models like the univariate case. The authors have examined the marginals and the minimum of the marginals, definitely they provide some indication about the goodness of fit of the proposed MOBW to the given data set. They have fitted $WE(1.7049, 5.443)$, $WE(1.7049, 4.3246)$ and $WE(1.7049, 6.3996)$ to X_1, X_2 and $\min\{X_1, X_2\}$ (using in fact that the distribution of marginals and the minimum of MOBW are Weibull ad by replacing the true values with their estimates). They have computed the Kolmogorov-Smirnov to fitted marginals and to the $\min\{X_1, X_2\}$, they are 0.123, 0.151 and 0.136 and the associated p-value are 0.631, 0.358, and 0.498, respectively. From the p-value , we cannot reject the hypotheses that X_1, X_1 and $\min\{X_1, X_2\}$ follow WE.

• For comparison purposes, they have reported the maximum likelihood estimates of α,λ_0,λ_1 and λ_2 and their confidence intervals given by Kundu and Dey (2009)(using EM algorithm):

unknown parameters	MLE	confidence interval
α	1.6954	[1.3284,2.0623]
λ_0	2.1927	[1.5001,2.8754]
λ_1	1.1192	[0.5411,1.6973]
λ_2	2.8852	[1.3023,4.4681]

Table 3.7: MLE and confidence interval of the unknown parameters

Important remarks

1. The estimators of α based on EM algorithm and the Bayesian method are very close to each other.

2. The length of the credible interval is smaller than the length of the confidence interval, which means that Bayesian method give more precision.

Conclusion

This dissertation mainely contains information related to the issues involved in Bayesian estimation of the Marshall-Olkin bivariate exponential distribution and some of its extensions in literature.

The simulations made using importance sampling show that the Bayesian estimates perform better than MLEs in the small sample case. Applying the same algorithm to a real data using the bivariate Weibull distribution of Marshall-Olkin, the results show that the lengths of confidence interval of the parameters based on Bayesian estimators are less than of confidence intervals of the parameters based on MLEs.

To improve this work, one can study the problem using many other various algorithms as Gibbs sampler, Lindley approximation.

Bibliography

[1] Arnold, B. C. (1968) Parameter estimation for a multivariate exponential distribution. J. Am. Statist. Ass., 63, 848-852.

[2] Balakrishnan, N., Lai, C.D., 2009. Continuous Bivariate Distributions. Springer, New York.

[3] Bemis, B. M., Bain, L. J. and Higgins, J. J. (1972) Estimation and hypothesis testing for the parameters of a bivariate exponential distribution. J. Am. Statist. Ass., 67, 927-929.

[4] Bhattacharyya, G. K. and Johnson, R. A. (1973) On a test of independence in a bivariate exponential distribution. J. Am. Statist. Ass., 68, 704-706.

[5] Calabria. R and Pulcini. G,(1994), An Engineering Approach to Bayes estimation for the Weibull distribution, Micron Electron Reliab, 34. 789-802.

[6] Chadli Assia, Talhi Hamida and Fellag Hocine, (2013), Comparison of the maximum likelihood and Bayes estimators for symmetric bivariate exponential distribution under different loss functions,AfrikaStatistika, 2316-090X , vol.8, pages 499- 514.

[7] Freund, J.E.: A bivariate extension of the exponential distribution. Journal of the American Statistical Association (1961).

[8] DeGroot .M.H, (1970), Optimal Statistical Decisions,´ed Mc-Graw-Hill.

[9] Downton, F.: Bivariate exponential distributions in reliability theory. Journal of the Royal Statistical Society, Series (1970).

[10] Galambos, J. and Kotz, S. (1978). characterizations of Probability Distributions, Springer-Verlag, Berlin-New Yok.

[11] Gumbel, E.J.: Bivariate exponential distributions. Journal of the American Statistical Association (1960).

[12] Hagen, E.W.: Common-mode/common-cause failure: A review. Annals of Nuclear Energy (1980).

[13] Jeevanand, E.S.: Bayes estimation of reliability under stress strength model for the Marshall and Olkin bivariate exponential distribution. IAPQR Trans.(1998).

[14] Klefsjo, B.: NBU and NBUE survival under the Marshall and Olkin shock model. IAPQR Trans. (1982).

[15] Kundu, D., and A. K. Dey. 2009. Estimating the parameters of the Marshall-Olkin bivariate Weibull distribution by EM algorithm. Computational Statistics and Data Analysis.

[16] Kundu, D., and A. K. Gupta. 2013. Bayes estimation for the Marshall-Olkin bivariate Weibull distribution. Computational Statistics and Data Analysis 57:271-81.

[17] Marshall, A.W., Olkin, I.: A multivariate exponential distribution. Journal of the American Statistical Association (1967).

[18] Meintanis, G., 2007. Test of fit for Marshall-Olkin distributions with applications. Journal of Statistical Planning and Inference 137, 3954-3963

[19] Muliere, P., Scarsini, M.: Characterization of a Marshall and Olkin type class of distributions. Annals of the Institute of Statistical Mathematics (1987).

[20] Nadarajah, S.: Exact distributions of XY for some bivariate exponential distributions. Statistics 40,(2006).

[21] Nadarajah, S., Kotz, S.: Sums and ratios for Marshall and Olkin's bivariate exponential distribution. Math. Sci. 30,(2005).

[22] Nadarajah, S., Zografos, K.: Expressions for Renyi and Shannon entropies for bivariate distributions. Information Sciences 170, (2005).

[23] Osaki, S., Yamada, S., Hishitani, J.: Availability theory for two-unit nonindependent series systems subject to shut-off rules. Reliab. Eng. Syst. Saf. 25,(1989).

[24] Pena, E.A., Gupta, A.K., 1990. Bayes estimation for the Marshall-Olkin exponential distribution. Journal of the Royal Statistical Society. Series B. Methodology 52, 379-389

[25] Proschan, F., Sullo, P.: Estimating the parameters of a bivariate exponential distribution in several sampling situations. In: Reliability and Biometry: Statistical Analysis of Life Lengths, F. Proschan and R.J. Serfling (eds.), pp. . Society for Industrial and Applied Mathematics, Philadelphia (1974).

[26] Ryu, K.: An extension of Marshall and Olkin's bivariate exponential distribution. J. Am. Stat. Assoc. 88, (1993).

[27] Ramanarayanan, R., Subramanian, A.: A 2-unit cold standby system with Marshall and Olkin bivariate exponential life and repair times. IEEE Trans. Reliab. R-30, (1981).

[28] Rai, K., Van Ryzin, J.: Multihit models for bivariate quantal responses. Stat. Decis. 2, (1984).

[29] Varian. H.R, (1975), A Bayesian Approach to real estate assessment,Amsterdam, North Hol- land. 195-208.